



# UNIVERSIDAD DE LA RIOJA

## TRABAJO FIN DE ESTUDIOS

Título

Análisis de la variación molecular asociada a un QTL con efecto pleiotrópico en la fertilidad y el tamaño de la baya en la vid

Autor/es

PABLO VILLOSLADA BLANCO

Director/es

PABLO CARBONELL BEJERANO , JOSÉ MIGUEL MARTÍNEZ ZAPATER y  
CARMEN TENORIO RODRÍGUEZ ,

Facultad

Escuela de Máster y Doctorado de la Universidad de La Rioja

Titulación

Máster Universitario en Química y Biotecnología

Departamento

AGRICULTURA Y ALIMENTACIÓN

Curso académico

2017-18



***Análisis de la variación molecular asociada a un QTL con efecto pleiotrópico en la fertilidad y el tamaño de la baya en la vid***, de PABLO VILLOSLADA BLANCO

(publicada por la Universidad de La Rioja) se difunde bajo una Licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 3.0 Unported. Permisos que vayan más allá de lo cubierto por esta licencia pueden solicitarse a los titulares del copyright.

**Trabajo de Fin de Máster**

**Análisis de la variación  
molecular asociada a un QTL  
con efecto pleiotrópico en la  
fertilidad y el tamaño de la baya  
en la vid**

Autor:

*Pablo Villoslada Blanco*

Tutor/es: Pablo Carbonell, Jose Miguel Martínez y Carmen Tenorio

**MÁSTER:**

**Máster en Química y Biotecnología (760M)**

**Escuela de Máster y Doctorado**



**UNIVERSIDAD  
DE LA RIOJA**

**AÑO ACADÉMICO: 2017/2018**



Dar las gracias a todo el personal del ICVV que me han ayudado siempre que han podido y en especial a Pablo Carbonell Bejerano que ha estado ahí siempre que lo he necesitado. Y por supuesto agradecer a mi familia y a mi novia todo el apoyo que me brindan día a día.



# ÍNDICE

<b>1.</b>	<b>Introducción</b>	<b>9</b>
<b>1.1.</b>	<b>Bayas</b>	<b>9</b>
<b>1.2.</b>	<b>Domesticación, aumento de la producción y diversificación de la forma y el tamaño de la baya</b>	<b>10</b>
<b>1.3.</b>	<b>Factores que determinan la fertilidad en la vid</b>	<b>11</b>
<b>1.4.</b>	<b>Anatomía y desarrollo de la uva</b>	<b>12</b>
1.4.1.	La flor de la uva y la anatomía del gineceo	12
1.4.2.	Desarrollo pre-antesis del gineceo de la vid	13
1.4.3.	Anatomía de la uva	14
1.4.4.	Desarrollo post-antesis del fruto en la vid	15
<b>1.5.</b>	<b>Procesos del desarrollo que determinan la forma y el tamaño de la baya</b>	<b>17</b>
<b>1.6.</b>	<b>Regulación genética de la forma y el tamaño del fruto</b>	<b>18</b>
<b>1.7.</b>	<b>Antecedentes del laboratorio</b>	<b>19</b>
<b>2.</b>	<b>Objetivos</b>	<b>24</b>
<b>3.</b>	<b>Metodología</b>	<b>25</b>
<b>3.1.</b>	<b>Material vegetal de partida</b>	<b>25</b>
<b>3.2.</b>	<b>Análisis bioinformático de los alelos del gen candidato</b>	<b>25</b>
<b>3.3.</b>	<b>Extracciones de DNA</b>	<b>25</b>
<b>3.4.</b>	<b>Diseño de cebadores para PCR y qPCR</b>	<b>26</b>
<b>3.5.</b>	<b>PCR</b>	<b>29</b>
<b>3.6.</b>	<b>Secuenciación Sanger</b>	<b>29</b>
<b>3.7.</b>	<b>Análisis de los datos de RNAseq</b>	<b>29</b>
<b>3.8.</b>	<b>qPCR</b>	<b>29</b>

<b>4.</b>	<b>Resultados y discusión</b>	<b>30</b>
<b>4.1.</b>	<b>Caracterización de los alelos de <i>BCNT2</i> en el parental Crimson Seedless e identificación de diferencias entre ambos alelos</b>	<b>30</b>
<b>4.2.</b>	<b>Confirmación por PCR seguida de secuenciación de los polimorfismos identificados en la región candidata</b>	<b>37</b>
4.2.1.	Validación de los SNPs identificados en el UTR3 del gen <i>VIT_05s0020g00200</i>	37
4.2.2.	Validación de la delección del CDS2 y del SNP del UTR2 del gen <i>VIT_05s0020g00200</i>	40
<b>4.3.</b>	<b>Análisis de la secuencia de los transcritos de los genes <i>BCNT</i> en los datos de RNAseq</b>	<b>43</b>
<b>4.4.</b>	<b>Cuantificación del número de copias de los genes de la familia <i>BCNT</i></b>	<b>46</b>
<b>5.</b>	<b>Conclusiones</b>	<b>49</b>
<b>6.</b>	<b>Bibliografía</b>	<b>50</b>
<b>7.</b>	<b>Anexos</b>	<b>57</b>
<b>7.1.</b>	<b>Abreviaturas</b>	<b>57</b>



# RESUMEN

La fertilidad de la planta y el tamaño de la baya son caracteres que condicionan la producción vitícola y la calidad final de la uva. En estudios previos de genética cuantitativa en una progenie  $F_1$  procedente del cruzamiento entre las variedades de uva de mesa Red Globe y Crimson Seedless, colocalizando en la misma región del cromosoma 5, se identificaron QTLs de efecto mayor para la fertilidad de las yemas, el tamaño y la forma de la baya. Análisis transcriptómicos globales en la progenie identificaron un gen candidato en esta región que se sobreexpresa tanto en yemas de individuos con baja fertilidad como en flores y frutos en desarrollo de individuos con baya grande y alargada.

En este TFM se han identificado polimorfismos entre los alelos del gen mediante el análisis bioinformático de datos de secuenciación masiva de BACs de la región cromosómica en cuestión en Crimson Seedless y se ha confirmado alguno de ellos mediante amplificación por PCR y secuenciación por Sanger en los progenitores de la progenie y en individuos  $F_1$  seleccionados por su genotipo en la región del QTL y su variación extrema para los fenotipos estudiados. Los resultados de esta secuenciación también identificaron el alelo ligado a la mutación. Mientras que estudios de PCR cuantitativa descartaron la posibilidad de que la sobreexpresión se deba a duplicaciones del gen en este alelo, mediante distintos estudios de secuenciación (PacBio, Sanger y RNAseq) se identificaron mutaciones candidatas en el mismo que podrían ser responsables de su sobreexpresión. Estudios funcionales futuros podrían confirmar si alguna de estas mutaciones causa la sobreexpresión constitutiva del gen y si esto altera pleiotrópicamente el desarrollo reproductivo, afectando a la determinación floral y a la morfogénesis del fruto, dando lugar a la reducción de la fertilidad y al alargamiento de la baya.

Palabras clave: fertilidad, tamaño, forma, uva, QTLs, polimorfismo.

# ABSTRACT

Plant fertility and berry size are traits conditioning grape quality and viticulture yield. In previous quantitative genetic studies in a F<sub>1</sub> progeny derived of a cross between the table cultivars Red Globe and Crimson Seedless cross, major QTLs colocalizing in the same region at chromosome 5 were detected for bud fertility, berry size and berry shape. Global transcriptomic analyses in the progeny identified a gene in this region that is overexpressed both in buds of individuals with low fertility and in developing flowers and fruits of individuals with large and elongated berry.

In this TFM, the bioinformatic analysis of massive sequencing data of BACs from the chromosomal region identified polymorphisms distinguishing the two alleles of the gene in Crimson Seedless. PCR amplification and Sanger sequencing of Crimson Seedles and Red Globe progenitors and progeny individuals selected for their genotype in the QTL region and its extreme variation for the phenotypes studied confirmed the presence of some of the polymorphisms. Sequencing results also identified the allele linked to the mutation. While quantitative PCR assays ruled out the hypothesis of gene duplications in that allele as the cause of the overexpression, diverse sequencing studies (PacBio, Sanger and RNAseq) identified candidate misexpression mutations in that allele. Future functional analyses could confirm if any of these mutations lead to the constitutive overexpression of the gene and to pleiotropic alterations in reproductive development, including flower determination and fruit morphogenesis, finally ending in the reduction of fertility and the elongation of the berry.

Keywords: fertility, size, shape, grape, QTLs, polymorphism.

# 1. INTRODUCCIÓN Y ANTECEDENTES.

## 1.1. Bayas.

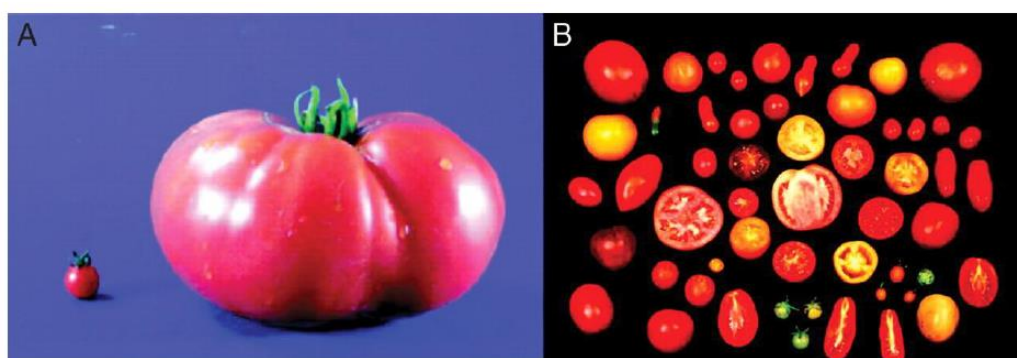
Los frutos son una innovación clave en las plantas angiospermas, ya que constituyen un receptáculo para el desarrollo de semillas, las protegen de daños por factores ambientales y/o depredadores y promueven su dispersión para la supervivencia de la especie (Dauelsberg *et al.*, 2011). Además, los frutos son una importante fuente nutricional y culinaria para el ser humano. Teniendo en cuenta la importancia económica y nutricional de los frutos frescos como componentes esenciales de las dietas animales y humanas, se requieren investigaciones considerables para mejorar la producción y la calidad de los frutos (Spjut, 1994).

La vid (*Vitis spp*, con la mayoría de cultivares pertenecientes a *Vitis vinifera* subsp. *vinifera*) es un cultivo de clima templado, una de las frutas más cultivadas a lo largo del mundo, cubriendo una superficie de 7,5 millones de hectáreas en 2016 y produciendo 75,8 millones de toneladas de bayas al año (OIV, 2018). Las uvas se clasifican dentro del grupo de las bayas, consistiendo en uno o más carpelos con una o más semillas y paredes ováricas carnales (Spjut, 1994). Son producidas a partir de un único ovario (Coombe, 1976) y son consideradas un buen modelo para los estudios sobre el desarrollo y la genética de los frutos (Dauelsberg *et al.*, 2011).

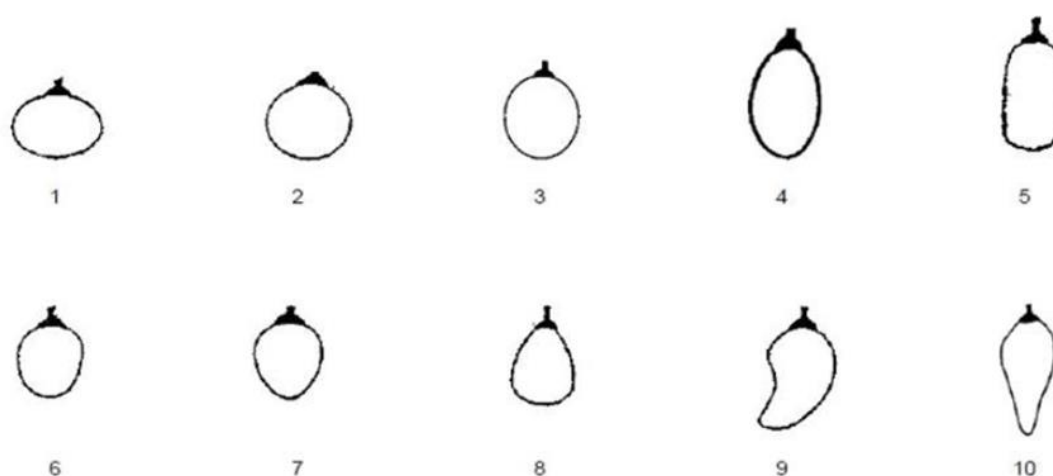
En la vid, la calidad de la baya es un factor determinante para la producción de vino y de uva de mesa. Esta depende de la interacción de parámetros genéticos (genotipo) y ambientales (suelo, clima, prácticas vitivinícolas, etc.) que determinan las características del desarrollo de la baya (tiempo de floración, fertilidad, dureza, forma, tamaño, etc.) y la composición de la misma (azúcares, ácidos orgánicos, metabolitos secundarios, etc.) (Dai *et al.*, 2011; Keller, 2015).

## 1.2. Domesticación, aumento de la producción y diversificación de la forma y el tamaño de la baya.

La domesticación es el resultado de un proceso de selección que busca incrementar la adaptación de las plantas (y los animales) al cultivo (y a la crianza) y a la utilización por parte de los seres humanos (Gepts, 2010). Las líneas de cultivo domesticadas que fueron seleccionadas para mejorar las características de los frutos normalmente llevan a frutos con mayor variabilidad en la forma, el tamaño y el color que sus parientes silvestres (Grandillo, Ku and Tanksley, 1999; Paris *et al.*, 2003). Por ejemplo, para el tomate existe una amplia variedad de tamaños y de formas (Tanksley, 2004) (**Figura 1**), de manera similar a lo que ocurre en la uva, donde existen desde formas esféricas a cilíndricas y hasta con aspecto de cuerno (**Figura 2**).



**Figura 1.** Ampla variación en la forma y el tamaño del tomate. **A)** El tomate gigante “Giant Heirloom” (derecha) y el tomate silvestre *L. pimpinellifolium* (izquierda). **B)** Rango de variación en el tamaño y la forma del tomate (Tanksley, 2004)



**Figura 2.** Tipos de uva de acuerdo a los descriptores OIV.

Los beneficios de incrementar el tamaño de los frutos para los humanos son claros: se requiere menos tiempo para obtener una determinada masa de fruto y se incrementa la producción por planta y por unidad de área (Tanksley, 2004).

La demanda por parte de la agricultura de incrementar el tamaño de los frutos ha contribuido a la evolución del tamaño de los frutos en diferentes cultivos. Es posible que la selección de frutos más grandes haya causado de manera indirecta un incremento en la variedad fenotípica del tamaño de los frutos. De alguna forma el incremento en el tamaño de los frutos podría haber permitido la expresión fenotípica de alelos “escondidos” con un efecto poco visible en los frutos pequeños (Tanksley, 2004). Además, las formas y colores novedosos resultan atractivos para los humanos y se han seleccionado tradicionalmente en la agricultura. Estos caracteres se persiguen también como marca de exclusividad del producto en programas de mejora de uva de mesa más recientes (**Figura 2**).

Los consumidores de uva de mesa prefieren uvas grandes sin semillas (Ebadi, Moghadam and Fatahi, 2009), mientras que, debido a que los compuestos aromáticos y fenólicos son más abundantes en la piel, las uvas pequeñas y esféricas (que tienen la mayor proporción de superficie de piel frente a pulpa) y con pocas semillas permiten obtener vinos de mayor calidad.

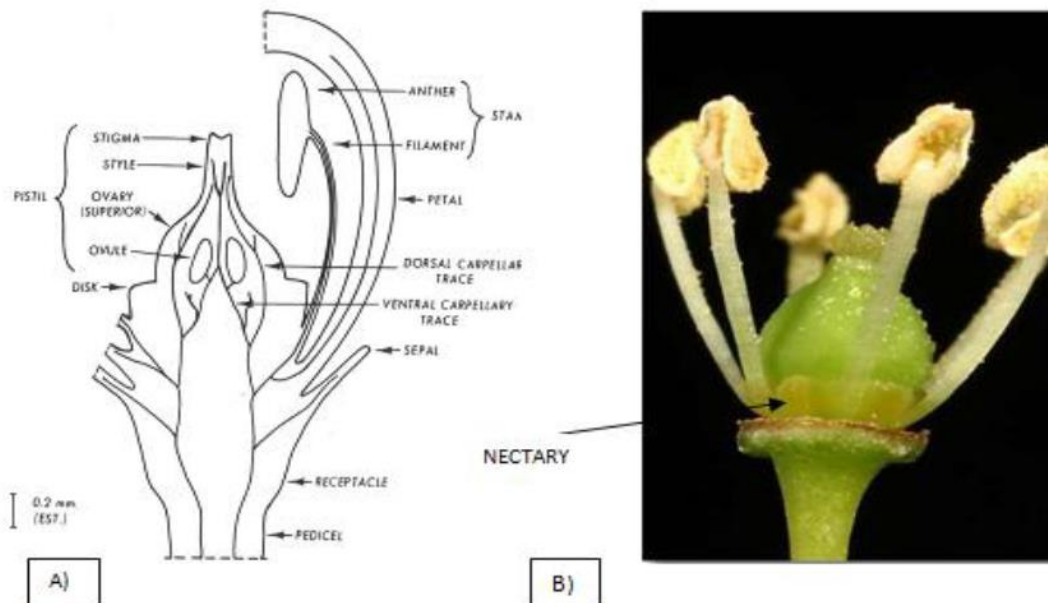
### **1.3. Factores que determinan la fertilidad en la vid.**

La fertilidad en la vid se mide en función del número de racimos por pámpano y depende de los meristemas laterales que se diferencian en brotes latentes (yemas) en el primer estadio (Sánchez and Dokoozlian, 2005). Estas yemas se desarrollan en la primavera y verano del primer año del ciclo y entran en un estado de latencia hasta que en la primavera siguiente dan lugar a inflorescencia o zarcillo, siendo esto un determinante clave en la fertilidad y el rendimiento (Carmona *et al.*, 2008). Es sabido que las uvas de vino son más fértiles que las uvas de mesa. Sin embargo, la fertilidad alta es un rasgo que se persigue en las uvas de mesa, mientras que es problemática en las uvas de vino, ya que se requieren rendimientos bajos para aumentar la calidad del vino, lo que incrementa los costes de producción (Sánchez and Dokoozlian, 2005).

## 1.4. Anatomía y desarrollo de la uva.

### 1.4.1. La flor de la uva y la anatomía del gineceo.

La uva se desarrolla a partir del gineceo (el órgano reproductor femenino), que en la vid se trata de un pistilo bicarpelar que se localiza en la espira interna de las flores de uva hermafroditas y femeninas (Pratt, 1971). El tamaño y la forma final de la baya vienen determinados por procesos que tienen lugar durante el desarrollo pre-antesis del pistilo y el desarrollo posterior de la fruta. Muchos cultivares de vid utilizados para la producción de uva presentan inflorescencia con flores hermafroditas (flores con estambres y pistilo), mientras que la vid silvestre (*Vitis vinifera* subsp. *silvestris*) presenta flores masculinas (flores con estambres) y femeninas (flores con pistilo) en distintos individuos (Pratt, 1971). Las flores hermafroditas de las vides de cultivo se seleccionaron como un carácter de domesticación para asegurar el cuajado de los frutos y la producción. Están compuestas por sépalos, pétalos, estambres y carpelos dispuestos en anillos concéntricos del exterior al interior (**Figura 3A**). En concreto, aparecen cinco sépalos que forman el cáliz y son la base de la flor protegiendo al resto durante los primeros estadios del desarrollo y cinco pétalos que forman una capa protectora sobre los órganos reproductivos y que crecen sobre el cáliz. En las flores de vid estos pétalos están unidos por células epidérmicas y forman la caliptra, la cual se cae durante la antesis, el momento en el que se abren las anteras y ocurre la polinización (Vasconcelos *et al.*, 2009). También aparecen cinco estambres que forman el androceo y un pistilo con un ovario superior que forma el gineceo (Dokoozlian, 2000). El pistilo está formado por la fusión de dos carpelos separados por un tabique que divide el ovario y la uva en dos lóbulos, cada uno conteniendo dos óvulos (Pratt, 1971). Por último, en la base de la flor aparecen cinco glándulas denominadas nectarios que sirven para atraer a los insectos polinizadores (Dokoozlian, 2000) (**Figura 3B**).



**Figura 3. A)** Diagrama de una sección longitudinal de una flor hermafrodita de vid antes de la antesis. **B)** Foto de una flor hermafrodita de vid tras la antesis.

#### 1.4.2. *Desarrollo pre-antesis del gineceo de la vid.*

La fase del desarrollo del gineceo en la vid corresponde a la morfogénesis y crecimiento del pistilo y los óvulos, desde la iniciación de la flor a la doble fecundación que tiene lugar en los óvulos (Gillaspy, Ben-David and Gruissem, 1993; Tanksley, 2004). Se pueden distinguir tres fases en la diferenciación del ovario: la formación del anillo meristémico, la proliferación celular por divisiones anticlinales y la maduración, caracterizada por las divisiones celulares periclinales y la diferenciación (Considine and Knox, 1979).

El primordio del pistilo se forma a partir de dos centros de crecimiento en el disco central de la flor (Considine and Knox, 1979). Cinco semanas antes de la antesis el pistilo se ha expandido lateral y verticalmente y ha dado lugar a cuatro óvulos. El tabique entre los carpelos ha aumentado de grosor por expansión y vacuolización de las células parenquimáticas. Posteriormente, cuatro semanas antes de la antesis se forma el estilo y los estigmas mediante divisiones periclinales del meristemo del pistilo (Considine and Knox, 1979).

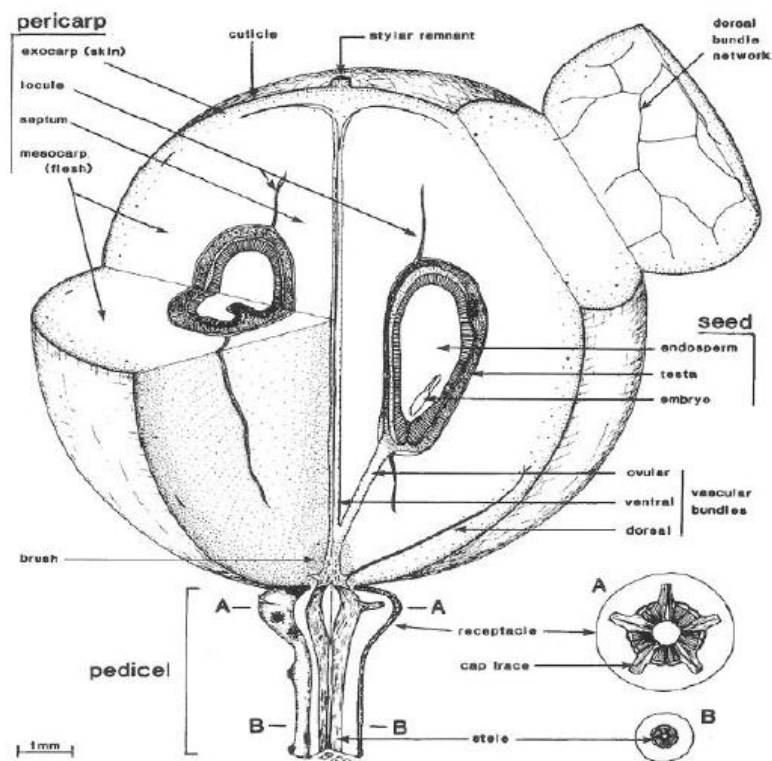
La maduración del pistilo es un periodo de intensa actividad que comienza tres semanas antes de la antesis. En esta fase se produce el crecimiento del estilo y los estigmas mediante divisiones periclinales y elongaciones axiales y también tiene lugar la meiosis en la megaspora (Considine and Knox, 1979).

Por último, tienen lugar divisiones periclinales, que en el parénquima ovárico van a determinar el número de células disponibles para el desarrollo del fruto. Es decir, el tamaño del fruto va a venir determinado por el número de células formadas durante esta fase del crecimiento, mientras que la forma va a venir determinada por el ratio entre las divisiones anticlinales que tienen lugar en los planos longitudinal y meridional (Considine and Knox, 1979).

#### 1.4.3. *Anatomía de la uva.*

Los tejidos de la baya derivan directamente de la pared del ovario y comprenden, del exterior al interior, la epidermis externa o exocarpo, el mesocarpo con las células parenquimáticas externas e internas, y la epidermis interna o endocarpo (Pratt, 1971) (**Figura 4**). Las células del exocarpo acumulan polifenoles en sus vacuolas, los cuales son importantes componentes del vino. Además, las vacuolas de las células epidérmicas contienen antocianinas responsables del color en las bayas rojas o negras (Coombe, 1976; Ollat *et al.*, 2002). Por otra parte, las células en el mesocarpo constituyen la carne de la baya y están especializadas en la acumulación de azúcares (principalmente glucosa y fructosa), ácidos orgánicos (principalmente ácido tartárico y ácido málico) y agua (Pratt, 1971).





**Figura 4.** Estructura de una uva madura.

#### 1.4.4. *Desarrollo post-antesis del fruto de la vid.*

Aproximadamente dos semanas después de que los óvulos se hayan formado y de cinco a diez semanas después de la brotación de las yemas de invierno (dependiendo del genotipo de la planta y las condiciones ambientales), la antesis marca el comienzo de la floración, exponiendo los órganos masculinos y femeninos (Keller, 2015).

Cuando las anteras se abren se libera el polen. La transferencia de los granos de polen (microgametofitos) de las anteras al estigma se denomina polinización y es seguida de la abscisión de los estambres (Keller, 2015). En los granos de polen adheridos al estigma se induce la germinación del tubo polínico. Los gametos masculino y femenino (células sexuales) se encuentran cuando el tubo polínico penetra y crece a través del estilo, alcanza el saco embrionario y suelta las células espermáticas. La fecundación (doble en este caso) tiene lugar cuando las dos células espermáticas se fusionan en el saco embrionario del óvulo con la célula madre y con los dos núcleos polares, respectivamente. En condiciones normales ocurre dos o tres días después de la polinización (Dokoozlian, 2000).

En las uvas, el cuajado del fruto se define cuando el diámetro de la baya está entre 1,6 y 3,2 mm. Los ovarios que no van a desarrollar fruto se caen en esta etapa, lo que es crítico para definir el número final de frutos que se van a desarrollar en el racimo (Dokoozlian, 2000).

En la uva, los procesos desde la fecundación hasta el estado de fruto están bajo el control de hormonas del desarrollo (auxinas, citoquinas y giberelinas) que promueven la división celular y la expansión (Coombe, 1960; Gillaspy, Ben-David and Gruissem, 1993; Zepeda *et al.*, 2006; Dauelsberg *et al.*, 2011). El cuajado del fruto está controlado por las auxinas y las giberelinas, las cuales inducen el desarrollo de los carpelos y juegan un papel clave en el inicio de la transformación del ovario al fruto (Coombe, 1960; Gillaspy, Ben-David and Gruissem, 1993; Ozga and Reinecke, 2003; Dauelsberg *et al.*, 2011). Además, los factores climáticos como la temperatura, la lluvia y la humedad juegan un papel clave en el cuajado y el desarrollo del fruto (Dokoozlian, 2000).

El desarrollo de la uva es un proceso dinámico que implica una serie de complejos cambios moleculares y bioquímicos divididos en tres grandes fases (Conde *et al.*, 2007; Deluc *et al.*, 2007). Después de la floración, durante la fase I, el crecimiento de la baya está basado en un número reducido de divisiones mitóticas y en expansiones celulares. Aproximadamente de cuatro a seis semanas post-antesis, las divisiones mitóticas cesan y solo se mantienen las expansiones celulares (Coombe, 1976). Al final de la fase I, todos los tejidos que componen las semillas están formados. La fase II corresponde al periodo de latencia, durante el cual las expansiones celulares cesan y los azúcares comienzan a acumularse. El final de la fase II y el inicio de la fase III marcan el envero o *véraison*, el inicio de la madurez. La fase III está caracterizada por un aumento masivo de varios componentes, principalmente glucosa y fructosa, los cuales se acumulan en las vacuolas de las células carnosas. Además, el crecimiento del fruto continua mediante expansión celular (Coombe, 1976; Ojeda *et al.*, 1999).

La madurez de la uva está controlada por numerosos factores. La duración de cada fase de desarrollo del fruto y las características finales de la baya (incluida la composición, la forma y el tamaño) dependen de la variedad, pero también pueden ser modificadas por factores ambientales como la intensidad de la luz, la temperatura, el suelo o la disponibilidad de nutrientes (Keller, 2015).

### **1.5. Procesos del desarrollo que determinan la forma y el tamaño del fruto.**

Los procesos del desarrollo involucrados en el control de la forma de la baya en la vid no se conocen claramente todavía. La forma y el tamaño de la baya se establecen principalmente durante la etapa de intenso crecimiento que ocurre en la primera fase de desarrollo de la baya antes del envero (Ojeda *et al.*, 1999; Houel *et al.*, 2013). Además, se ha determinado que el número de divisiones celulares que tiene lugar durante el desarrollo pre-antesis del pistilo y la magnitud de las expansiones celulares durante el desarrollo de la baya después de la antesis están implicadas en la variación del peso de la pulpa entre las distintas variedades de vid (Coombe, 1972, 1976).

Por otro lado, se ha demostrado la correlación entre el número de células parenquimáticas en la pared ovárica durante la antesis y la forma del fruto en el tomate, la oliva y el melón (Périn *et al.*, 2002; Bertin *et al.*, 2003; Cheniclet *et al.*, 2005; Baldet *et al.*, 2006; Cong, Barrero and Tanksley, 2008; Rosati *et al.*, 2011). En el tomate, el gen *OVATE* promueve la forma de pera induciendo un aumento en la división celular en la región proximal del ovario en desarrollo (Monforte *et al.*, 2014), el gen *SUN* induce la forma elongada del tomate al inducir elongación celular en ambos polos (Van der Knaap and Tanksley, 2003) y el gen *FRUIT SHAPE 8.1* es responsable de la apariencia compacta de los tomates procesados (Grandillo, Ku and Tanksley, 1999).

También se ha encontrado una fuerte correlación entre la diversidad en la forma y el tamaño del tomate (Tanksley, 2004) que podría ser debida a un locus con efecto pleiotrópico o a una selección paralela pero independiente de ambos rasgos para satisfacer las demandas de los consumidores (Tanksley, 2004).

Otro factor que afecta al tamaño y la forma del tomate es el número de carpelos (Cong, Barrero and Tanksley, 2008). Sin embargo, en la vid este rasgo es mucho menos variable (Houel *et al.*, 2013) y tiene más influencia la localización de la uva en el racimo, el número de uvas por racimo y la relación fuente/sumidero de la vid (Ollat *et al.*, 2002; Dai *et al.*, 2011).

## **1.6. Regulación genética de la forma y el tamaño del fruto.**

Los mecanismos genéticos que controlan la forma y el tamaño de la uva permanecen desconocidos, pero han sido estudiados en otras especies con fruto carnoso como el tomate, en el que se han encontrado varios Quantitative Trait Loci (QTLs) responsables de la forma del fruto.

Los cambios en la forma del fruto en el tomate se deben a mutaciones en cuatro genes: *SUN*, *OVATE*, *LOCULE NUMBER* Y *FASCIATED* (Rodríguez *et al.*, 2011). *LOCULE NUMBER* y *FASCIATED* afectan a la forma y el tamaño del fruto mediante la variación en el número de carpelos de la flor (Barrero and Tanksley, 2004). *FASCIATED* codifica un factor de transcripción (FT) tipo YABBY cuya falta de expresión da lugar a tomates con un gran número de lóculos, muy frecuentes en los cultivos modernos (Cong, Barrero and Tanksley, 2008). *OVATE* codifica una proteína de la familia Ovate Family Protein (OFP) y se piensa que regula negativamente la transcripción de los genes diana (Liu *et al.*, 2002; Hackbusch *et al.*, 2005). La forma de pera es debida a una mutación puntual recesiva que genera un codón de terminación en el segundo exón y da lugar a una proteína truncada. Por otro lado, la forma de fruto elongada debida al gen *SUN* se relaciona con una duplicación intercromosómica (Lozano *et al.*, 2009).

El análisis genético del tamaño y la forma de la baya en la vid es más complicado que en otras especies debido a su largo ciclo biológico, a su elevada heterocigosidad y a la gran superficie necesaria para su cultivo (Doligez *et al.*, 2002; Cabezas *et al.*, 2006; Mejía *et al.*, 2007; Le Cunff *et al.*, 2008). Muchas de las variantes genéticas naturales están almacenadas en los centros de germoplasma y se están haciendo grandes esfuerzos para caracterizarlas (Le Cunff *et al.*, 2008) mediante estudios de asociación génica (Fournier-Level *et al.*, 2009; Fernandez *et al.*, 2013) y análisis de QTLs (Doligez *et al.*, 2002; Cabezas *et al.*, 2006; Le Cunff *et al.*, 2008).

Los análisis de QTLs identifican regiones del genoma que contribuyen a la variación de un carácter concreto y son el punto de partida para estrechar esa región a un punto donde los efectos puedan ser asociados a genes específicos albergados en esas regiones y en última estancia, a mutaciones o polimorfismos causales en dichos genes (Le Cunff *et al.*, 2008). En otros casos, la información sobre la regulación del tamaño de la uva procede del análisis de variantes espontáneas o de estudios moleculares (Fernandez *et al.*, 2013).

## **1.7. Antecedentes del laboratorio.**

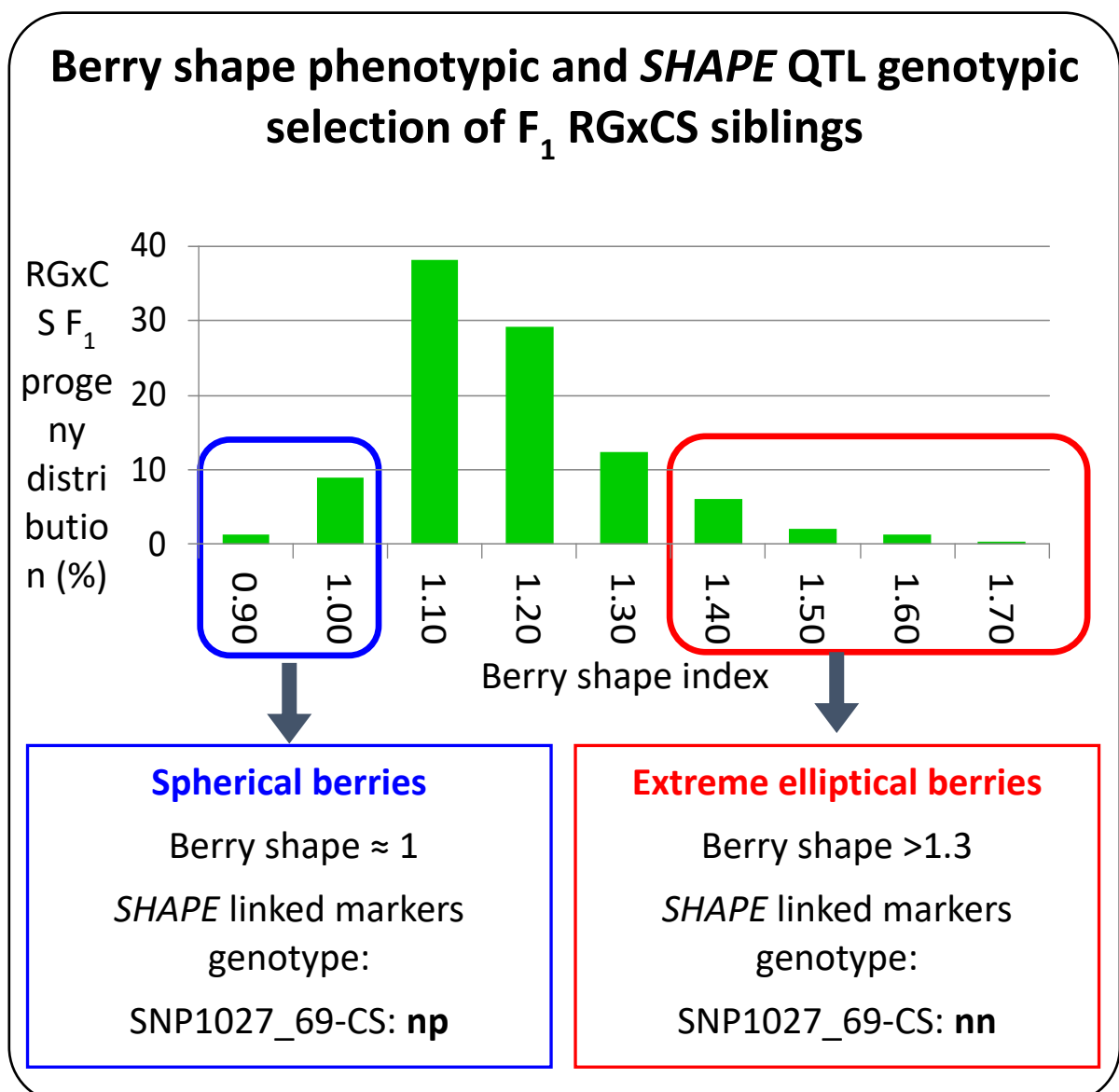
En trabajos anteriores del grupo se llevó a cabo un análisis de QTLs en una progenie F<sub>1</sub> resultado de un cruzamiento entre las variedades de uva de mesa Red Globe (RG) y Crimson Seedles (CS) (Royo *et al.*, 2018). Esta población segrega para varios rasgos relevantes para los que se analizó el fenotipo durante tres años (2007, 2008 y 2009). Entre otros caracteres se analizó la fertilidad y el tamaño y la forma de la baya. En este estudio se identificaron seis QTLs de efecto mayor en la longitud de la baya en los grupos de ligamiento (LG) 5, 6, 8, 14, 18 y 19, respectivamente (SHAPE, BL6a, BL8a, BL14a, SDI y BL19), ocho QTLs de efecto mayor sobre el diámetro de la baya en los LG 1, 5, 6, 8, 12, 14 y 18 (BD1a, BD1b, SHAPE, BD6a, BD8b, BD12a, BD14a y SDI) y nueve QTL de efecto mayor sobre la variación en la forma de la baya localizados los LG 1, 5, 7, 8, 10, 11, 14, 18 y 19 (BS1b, SHAPE, BS7, BS8a, BS10, BS11, BS14a, BS18a y BS19). Finalmente, para el carácter fertilidad, se identificaron cuatro QTL de efecto mayor en los LG 5, 10, 14 y 18, respectivamente (FERTILITY, IF10, IF14 y IF18)

Los QTL *FERTILITY* y *SHAPE* co-localizan en la misma región del cromosoma 5 que tiene efecto mayor en la variación de los cuatro rasgos mencionados anteriormente. Dada la naturaleza de efecto mayor y pleiotrópico del QTL *FERTILITY/SHAPE*, para identificar genes candidatos subyacentes, se llevaron a cabo varios análisis transcriptómicos comparando la expresión génica entre individuos de la progenie F<sub>1</sub> utilizando micomatrices NimbleGen de expresión global de vid. Por un lado, se compararon individuos F<sub>1</sub> con fenotipo extremo en el índice forma de baya (definido como la longitud de la baya dividida entre el diámetro ecuatorial de la baya). Estos individuos también fueron seleccionados por su genotipo en el QTL *FERTILITY/SHAPE* de modo que presentasen la combinación alélica característica del extremo de la distribución correspondiente, baya alargada o baya esférica (**Figura 5**). De esta forma se perseguía caracterizar la base molecular del efecto del QTL *SHAPE* en la forma de la baya e identificar genes candidatos responsables de dicho efecto. Para esta comparación transcriptómica se utilizó RNA de flores en desarrollo ya que estudios histológicos y morfométricos previos del grupo mostraron que ya hay diferencias en la forma del pistilo entre los individuos de cada cola y que la diferencia va incrementando desde este momento hasta el envero. Este estudio indicó que había únicamente 55 genes con expresión diferencial entre los individuos de baya alargada y al comparar estos resultados con los obtenidos en un experimento independiente similar sobre la misma progenie, pero realizado en un estadio pre-antesis de flor más temprano, se reveló que solo 6 genes coincidían. De estos 6 genes, *VIT\_05s0020g00200* y *VIT\_05s0020g00210*, presentaban los mayores valores de sobreexpresión en los individuos con baya alargada y se localizan en el intervalo de confianza del QTL *SHAPE*, por lo que se tomaron como genes candidatos. Aparte de estos dos genes también se consideraron *VIT\_05s0020g00080*, *VIT\_05s0020g00090* y *VIT\_05s0020g00360*, ya que presentaban un alto grado de homología con los dos genes anteriores, todos se localizan en el intervalo de confianza del QTL en el cromosoma 5 y además *VIT\_05s0020g00080* y *VIT\_05s0020g00090* también estaban sobreexpresados en los individuos con baya alargada, aunque en menor medida.

De modo similar, se realizó un análisis transcriptómico comparando individuos  $F_1$  de los extremos de la distribución del índice de fertilidad que fueron seleccionados también según la combinación alélica más frecuente en el QTL *FERTILITY/SHAPE* en el correspondiente extremo de la distribución. Para la comparación se utilizó RNA de yemas latentes en dos estadios, correspondientes al inicio de la floración y al cuajado, momentos en los que en la yema se produce la determinación floral y se inicia el desarrollo de los primordios de inflorescencias o zarcillos, respectivamente. En este estudio, considerando los dos estadios en conjunto, se detectaron 756 genes con expresión diferencial entre individuos de ambos extremos de la distribución de fertilidad. De nuevo, los genes *VIT\_05s0020g00200* y *VIT\_05s0020g00210* fueron los que mostraron los mayores niveles de sobreexpresión en ambos estadios de yema en individuos de baja fertilidad (que correlaciona con que son individuos que desarrollan bayas más alargadas). Los genes *VIT\_05s0020g00080* y *VIT\_05s0020g00090* también mostraron una sobreexpresión de menor magnitud en las yemas en individuos de baja fertilidad. Posteriormente, se llevó a cabo un análisis estructural de estos genes que mostró que *VIT\_05s0020g00080* y *VIT\_05s0020g00090* por un lado y *VIT\_05s0020g00200* y *VIT\_05s0020g00210* por otro, están contiguos en el genoma y que *VIT\_05s0020g00080* y *VIT\_05s0020g00360* presentan el dominio Bucentaurus (*BCNT*) (**Figura 6**). El dominio *BCNT* es característico de los genes de la familia *BCNT* que en otros organismos eucariotas, como *Saccharomyces cerevisiae* (levadura), *Drosophila melanogaster* (mosca) y *Homo sapiens* (humano), se ha visto que participan en el complejo multiproteico SWR1 implicado en la organización de la cromatina y en el control epigenético de la expresión génica mediante el reemplazamiento de histonas (Messina *et al.*, 2014, 2015, 2016, 2017; Sun and Luk, 2017). Por ello, en este trabajo *VIT\_05s0020g00080* y *VIT\_05s0020g00090* pasarán a denominarse locus *BCNT1*, *VIT\_05s0020g00200* y *VIT\_05s0020g00210* pasarán a denominarse locus *BCNT2* y *VIT\_05s0020g00360* pasará a denominarse locus *BCNT3*.

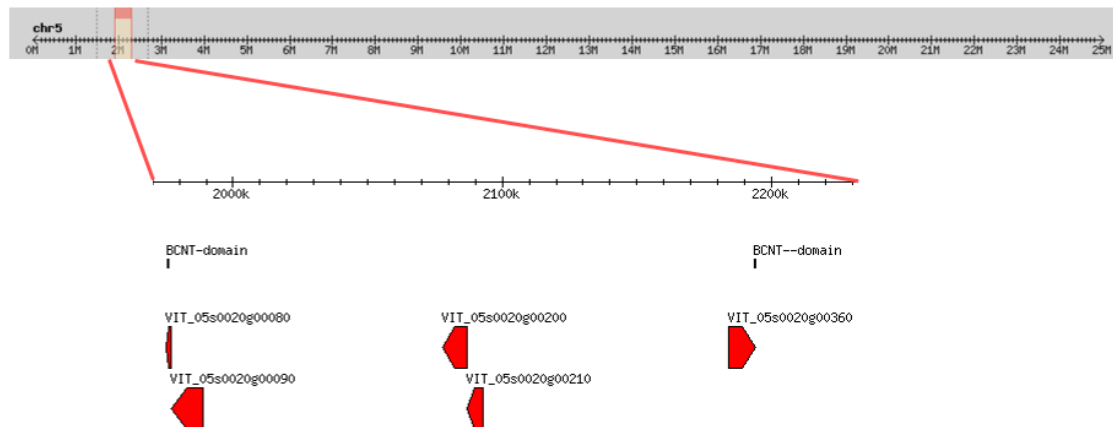


Finalmente, dada la elevada homología entre *BCNT1* y *BCNT2*, en estudios previos se evaluó si la sobreexpresión observada en los análisis de micromatrices para alguno de ellos podría ser debida a una hibridación cruzada con el otro miembro. Los resultados revelaron que muy posiblemente las diferencias en la expresión diferencial observadas para *BCNT1* eran debidas a una hibridación cruzada con transcritos de *BCNT2* (Borzecka, 2014), por lo que en este trabajo abordaremos los estudios centrándonos sobre todo en *BCNT2* como gen candidato del QTL *FERTILITY/SHAPE*.



**Figura 5.** Distribución de los individuos de la progenie F<sub>1</sub>. El cuadro azul representa los individuos con baya esférica y aparecen los marcadores del QTL *SHAPE* asociados. El cuadro rojo representa los individuos con baya esférica y aparecen los marcadores del QTL *SHAPE* asociados (Borzecka, 2014).





**Figura 6.** Esquema de la estructura de los locus *BCNT1*, *BCNT2* y *BCNT3* (Borzecka, 2014).

La hipótesis de partida apoya la existencia de una mutación en uno de los alelos del gen *BCNT2* de cada uno de los parentales (RG y CS) de la progenie estudiada. Esta mutación causaría la sobreexpresión constitutiva del gen, la cual estaría ligada a efectos pleiotrópicos en el desarrollo reproductivo de la vid, incluyendo una disminución de la inducción de la floración en las yemas y un aumento en el crecimiento de la baya, sobre todo en el sentido longitudinal. Para intentar contrastar la hipótesis, lo primero que se hizo fue tratar de obtener la secuencia de los dos alelos de *BCNT2* y sus regiones reguladoras en CS. Para ello se llevó a cabo una secuenciación masiva mediante la tecnología de PacBio de *bacterial artificial chromosomes* (BACs) de la región cromosómica en cuestión del parental CS.

## 2. OBJETIVOS.

El objetivo general de este trabajo es comprender el origen genético del QTL *FETILITY/SHAPE* mediante la identificación de las mutaciones responsables de la sobreexpresión del gen candidato *BCNT2*. La identificación del polimorfismo causal se podría explotar, en programas de mejora clásica o de edición génica, para aumentar el aumento del tamaño de la baya y la diversificación de su forma, así como para controlar los niveles de producción a partir de la fertilidad de las yemas. Con esto también se pretende comprender cuál es la función molecular de la familia de genes *BCNT* y su papel en el desarrollo de las plantas ya que todavía no se ha descrito. Para poder avanzar en estos objetivos generales, en el marco temporal que permite la duración del proyecto, se han planteado los siguientes objetivos parciales:

- 2.1.** Caracterización de los alelos del gen *BCNT2* en la variedad Crimson Seedless e identificación de polimorfismos de secuencia entre ambos a partir del ensamblaje de los datos de secuenciación masiva.
- 2.2.** Confirmación por PCR seguida de secuenciación Sanger de los polimorfismos identificados en *BCNT2* y asociación de los polimorfismos con los alelos del QTL. Los experimentos se centrarán en los *single nucleotide polymorphisms* (SNPs) en regiones codificantes que permitan identificar el alelo que esté sobreexpresado y en INDELs en regiones reguladoras que puedan explicar la sobreexpresión del alelo mutado.
- 2.3.** Cuantificación del número de copias de los genes de la familia *BCNT* en cada alelo para determinar su posible relación con la sobreexpresión.

### 3. METODOLOGÍA.

#### 3.1. Material vegetal de partida.

Una población  $F_1$  de uva de mesa (*Vitis vinifera*) fue generada mediante cruzamientos controlados llevados a cabo en 2003. Una planta de cada individuo  $F_1$  y tres plantas de cada progenitor fueron cultivadas en las mismas condiciones experimentales en los campos de la Sociedad Murciana de Investigación y Tecnología de Uva de Mesa (ITUM) (Royo *et al.*, 2018).

#### 3.2. Análisis bioinformático de los alelos del gen candidato.

En el laboratorio se había realizado la secuenciación mediante PacBio de dos BACs de cada alelo de la región del genoma que contiene el gen candidato *BCNT2*. Mediante análisis bioinformáticos en el grupo habían ensamblado varios *contigs* a partir de los *reads* largos de PacBio y se planteó una comparación de los mismos para caracterizar los posibles alelos de la región y del gen candidato. El análisis de la secuencia del gen candidato fue realizado principalmente mediante BLAST (Altschul *et al.*, 1990), una herramienta específicamente diseñada para buscar similitud con secuencias de ácidos nucleicos y proteínas depositadas en bases de datos y comparar dos secuencias de nucleótidos o aminoácidos entre sí. También se utilizó BLAT (Kent, 2002), una herramienta de alineamiento similar a BLAST (Altschul *et al.*, 1990), pero estructurada diferentemente de modo que permite identificar más fácilmente la similitud cuando hay alternancia de fragmentos con y sin similitud en una de las secuencias frente a la otra.

#### 3.3. Extracciones de DNA.

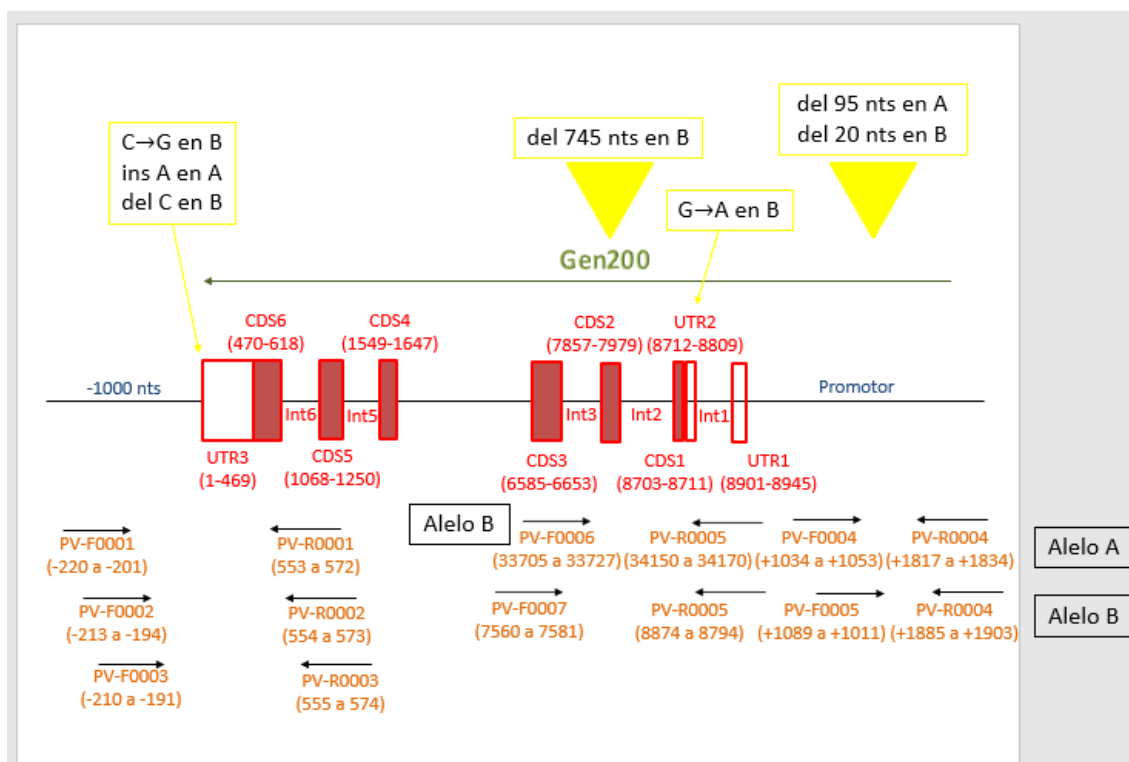
Para la extracción del DNA se utilizaron hojas del parental CS almacenadas a -80 °C y se siguió el protocolo de DNeasy® Plant Mini Kit (Qiagen) (*DNA extraction*, 2018). Para comprobar la calidad de las extracciones realizadas se llevó a cabo un análisis mediante Nanodrop (*Nanodrop*, 2018) en el que se cuantificó el contenido de DNA y se determinó su pureza en función de los ratios de absorbancia  $A_{260/280}$  y  $A_{260/230}$ .

### 3.4. Diseño de cebadores para PCR y qPCR.

Los cebadores fueron diseñados manualmente mediante el software OligoExplorer (Kuulasmaa, Javed and Muthusamy, 2004) que permite valorar las propiedades de los oligonucleótidos seleccionados. En la selección de parejas de cebadores se trató de que cumpliesen en la medida de lo posible los siguientes criterios:

- Tamaño del oligo: 18-25 nucleótidos. Diferencia entre cebadores de la misma pareja de PCR  $\leq 5$ .
- Temperatura de *melting* ( $T_m$ ): 58-61°C. Diferencia entre cebadores  $\leq 1^\circ\text{C}$ .
- %GC: 40-60%. Mejor si es próximo a 60%.
- Cola 3' GC: 40%. De los 5 últimos nucleótidos, un máximo de 2 G o C (o 3 si no hay dos piridinas consecutivas [TC ó CT]).
- Extremo 3': preferiblemente, que no sea T, porque tiene mayor tendencia a amplificación inespecífica.
- Repetición de nucleótidos: máximo de 4 (3 mejor) repeticiones consecutivas de un nucleótido o de una pareja de nucleótidos.
- Lo óptimo para PCR cuantitativa (qPCR) usando SYBR Green como fluorocromo es un contenido en C>G, porque el nucleótido G apaga la fluorescencia.
- Autoanillamiento y bucles: se descarta el cebador si se dan anillamientos con energía  $\geq -2$  kcal/mol en el extremo 3' del cebador y  $\geq -3$  kcal/mol en el extremo 5'.
- Anillamiento entre cebadores: se descarta la pareja de cebadores si se dan anillamientos entre ellos con energía  $\geq -5$  kcal/mol en el extremo 3' del cebador y  $\geq -6$  kcal/mol en el extremo 5'. En general, intentar que los oligos no aparezcan de ningún modo con  $T_m > 10^\circ\text{C}$ .
- FAMILIAS GÉNICAS: tratar de diseñar los oligos en la zona 3'UTR (*untranslated region*) para poder discriminar entre genes de la misma familia que suelen presentar mayor similitud en las regiones codificantes.

Una vez diseñados (**Figura 7, Tabla 1 y Tabla 2**) se determinó la especificidad de los cebadores mediante BLAST (*Grape BLAST*, 2018) sobre la secuencia genómica para testar que no amplifiquen en otra región.



**Figura 7.** Representación de la posición donde hibrida cada cebador y de los polimorfismos que se pretende analizar con cada pareja.

**Tabla 1.** Longitud prevista del producto de PCR con cada pareja de cebadores.

Forward	Reverse	Longitud
PV_F0001	PV_R0001	1354
PV_F0001	PV_R0002	1355
PV_F0001	PV_R0003	1356
PV_F0002	PV_R0001	1361
PV_F0002	PV_R0002	1362
PV_F0003	PV_R0002	1365
PV_F0003	PV_R0003	1366
Forward	Reverse	Longitud
PV_F0004	PV_R0004	801
PV_F0005	PV_R0004	815
Forward	Reverse	Longitud
PV_F0006	PV_R0005	466
PV_F0007	PV_R0005	1235
Forward	Reverse	Longitud
PV_F0008	PV_R0006	80
PV_F0009	PV_R0007	145
Forward	Reverse	Longitud
PV_F0010	PV_R0008	194
PV_F0011	PV_R0009	171

**Tabla 2.** Información de los cebadores sintetizados.

Nombre	Secuencia	T <sub>m</sub> (°C)	Gen	Posición 5'	Descripción
PV_F0001	TATTCGAGCCTTGGAACTC	59,2	1000ntsUpstreamVIT_05s0020g00200	220	Upper primer para amplificar los polimorfismos del UTR3
PV_F0002	TAACCCATATTCGAGCCTTG	59,9	1000ntsUpstreamVIT_05s0020g00200	213	Upper primer para amplificar los polimorfismos del UTR3
PV_F0003	ACATAACCCATATTCGAGCC	58,7	1000ntsUpstreamVIT_05s0020g00200	210	Upper primer para amplificar los polimorfismos del UTR3
PV_R0001	GCTTGTGATGCTTCAGAGG	61,0	VIT_05s0020g00200	572	Lower primer para amplificar los polimorfismos del UTR3
PV_R0002	AGCTTGTGATGCTTCAGAG	58,2	VIT_05s0020g00200	573	Lower primer para amplificar los polimorfismos del UTR3
PV_R0003	AAGCTTGTGATGCTTCAGA	58,6	VIT_05s0020g00200	574	Lower primer para amplificar los polimorfismos del UTR3
PV_F0004	ACCGTGGTAGGTGATAGGAC	58,7	PromotorVIT_05s0020g00200	1034	Upper primer para amplificar la delección del promotor del AleloA
PV_F0005	TGAGAAGTGACTTGTGTCAGTG	60,4	PromotorVIT_05s0020g00200	1089	Upper primer para amplificar la delección del promotor del AleloB
PV_R0004	AAAAATCGAGTTTCGCC	60,0	PromotorVIT_05s0020g00200	1834 AleloA 1903 AleloB	Lower primer para amplificar las delecciones del promotor del AleloA y del Alelo B
PV_F0006	TTCAATCTCAAGCTTAACAGTAA	56,0	VIT_05s0020g00200	33705 AleloB	Upper primer para amplificar la delección del Int3+CDS2+Int2 del AleloB
PV_F0007	GGTTGATGGTTCACCTTTGTC	58,3	VIT_05s0020g00200	7560	Upper primer para amplificar la región de la delección del Int3+CDS2+Int2
PV_R0005	TGATTGTTCAATTTGGTTTTG	58,4	VIT_05s0020g00200	34170 AleloB	Lower primer para amplificar la región de la delección del Int3+CDS2+Int2 y el SNP del UTR2
PV_F0008	TTCAACTTTTCTCTGCCTG	60,4	VIT_05s0020g00200	1068	Upper primer para cuantificar el número de copias del Gen200, del Gen80 y del Gen360 mediante qPCR
PV_R0006	GAGGTGAAGAGACTTGCTGC	60,1	VIT_05s0020g00200	1147	Lower primer para cuantificar el número de copias del Gen200, del Gen80 y del Gen360 mediante qPCR
PV_F0009	ATTAGCCTCGTTCAATTTC	59,2	VIT_05s0020g00200	173	Upper primer para cuantificar el número de copias del Gen200, del Gen80 y del Gen360 mediante qPCR
PV_R0007	AGGCCAGATATGAGAGAGGA	58,8	VIT_05s0020g00200	1147	Lower primer para cuantificar el número de copias del Gen200, del Gen80 y del Gen360 mediante qPCR
PV_F0010	CTGGACGACATGATGAACAG	60,1	VIT_17s0000g00150	271	Upper primer para amplificar el GenLEAFY como control para cuantificar el número de copias del Gen200
PV_R0008	GATGGTAACGCACCTTCTTG	60,5	VIT_17s0000g00150	464	Lower primer para amplificar el GenLEAFY como control para cuantificar el número de copias del Gen200
PV_F0011	CAAGAAGGTGCGTTACCATC	60,5	VIT_17s0000g00150	445	Upper primer para amplificar el GenLEAFY como control para cuantificar el número de copias del Gen200
PV_R0009	CTCGACAGTCTGCTTATG	60,8	VIT_17s0000g00150	615	Lower primer para amplificar el GenLEAFY como control para cuantificar el número de copias del Gen200
PV_F0012	GAGGGGAAGATACCAACA	58,9	VIT_05s0020g00200	1000	Upper primer para secuenciar los polimorfismos del UTR3

### 3.5. PCR.

Para llevar a cabo los ensayos de PCR, el correspondiente transcrito fue amplificado utilizando los cebadores de la **Tabla 2** y el kit MyTaq™ DNA Polymerase ('PCR', 2018). Para comprobar que el tamaño del producto de PCR era el esperado y que no se obtenían otras bandas debido a amplificación inespecífica, se llevó a cabo una electroforesis en un gel de agarosa.

### 3.6. Secuenciación Sanger.

Para llevar a cabo la secuenciación, el producto de amplificación resultante fue purificado mediante ExoSAP-IT (*ExoSAP-IT*, 2018) y fue enviado al servicio de secuenciación de electroforesis capilar Sanger de STAB VIDA (*Secuenciación Sanger*, 2018). Posteriormente, los resultados de las secuencias obtenidas fueron analizados mediante el software SeqScape ('SeqScape v2.6', 2018).

### 3.7. Análisis de los datos de RNAseq.

Para el análisis de los datos de RNAseq obtenidos en trabajos previos del grupo (Royo *et al.*, 2018) se utilizó el software IGV ('IGV\_2.4.8', 2018).

### 3.8. qPCR.

Para llevar a cabo los ensayos de qPCR, el correspondiente transcrito fue amplificando usando los cebadores de la **Tabla 2** y el kit Power SYBR® Green PCR Master Mix (*qPCR*, 2018). Para cada muestra se realizaron tres réplicas técnicas. El número de copias de los genes *BCNT* fue determinado en función del número de ciclos de amplificación necesarios para alcanzar el umbral fijado en la fase exponencial de la reacción de amplificación ( $C_t$ ).

## 4. RESULTADOS Y DISCUSIÓN.

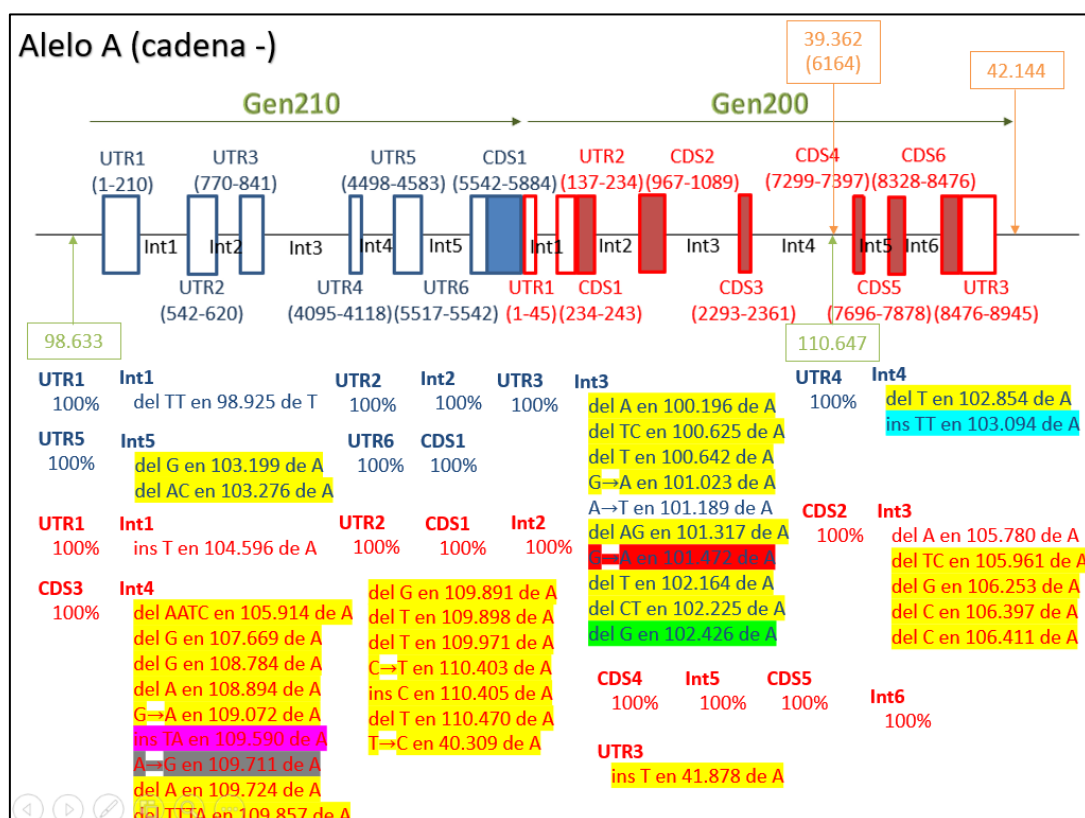
### 4.1. Caracterización de los alelos de *BCNT2* en el parental Crimson Seedless e identificación de diferencias entre ambos alelos.

Para identificar polimorfismos entre los alelos del gen candidato se procedió al análisis bioinformático del ensamblaje obtenido a partir de datos de secuenciación masiva de BACs de la región cromosómica en cuestión en la variedad CS. En este ensamblaje se identificaron contigs de dos alelos distintos, A y B, que fueron caracterizados en este trabajo como se describe a continuación.

En primer lugar, se identificaron polimorfismos tanto del Alelo A (**Figura 8**) como del Alelo B (**Figura 9**) frente al genoma de referencia de referencia de la vid. En concreto las comparaciones se realizaron frente a la versión 12X.0 del genoma de referencia construido a partir del genotipo homocigótico PN40024 derivado de varios ciclos de autofecundación de Pinot Noir (Jaillon *et al.*, 2007). Para ello, se llevó a cabo un BLAST (BLAST, 2018) de la secuencia del locus *BCNT2* (genes *VIT\_05s0020g00200* y *VIT\_05s0020g00210*, a partir de ahora Gen 200 y Gen 210, respectivamente) del genoma de referencia con respecto a los contigs del ensamblaje de CS. Mediante los alineamientos BLAST se determinó que el contig01 y el contig04 deben corresponder a alelos diferentes del gen en CS que se denominaron A y B, respectivamente. Se confirmó mediante un BLAST (BLAST, 2018) que el Gen 200 y el Gen 210 aparecían contiguos tanto en el genoma de referencia como en los alelos A y B. Sin embargo, los contigs 01 y 04 no contienen la secuencia completa del Gen 210. En concreto, el Gen 200 no alineaba con el contig01 a partir del nucleótido (nt) 6.136 en el caso del Alelo A, lo mismo ocurría a partir del nt 602 en el contig04 en el caso del Alelo B.



Posteriormente, para intentar resolver las partes que no alineaban se llevó a cabo un BLAT (BLAT, 2018) de los contigs 01 y 04 frente al genoma de referencia y una vez ensamblada la secuencia del Alelo B para el locus *BCNT2* completo se pudo completar su caracterización. Por otra parte, para completar el Alelo A del locus *BCNT2* tuvimos que usar otro contig (denominado contig377box).



**Figura 8.** Caracterización del Alelo A de *BCNT2*. Se encuentran marcadas las posiciones del contig01 y del contig377box que alinean con el genoma de referencia y la posición en la que el Gen 200 deja de alinear con el contig01 y pasa a alinear con el contig377box. Aparecen identificadas las *untranslated regions* (UTRs), las secuencias codificantes (CDSs) y los intrones (Int) según la anotación del genoma de referencia. Para cada fragmento están anotadas los polimorfismos identificados entre el Alelo A y el genoma de referencia. En fosforito se presentan los polimorfismos entre el Alelo A y el genoma de referencia que también implican una diferencia entre el Alelo A y el Alelo B. En fosforito con un color diferente al amarillo están anotados cada uno de los polimorfismos del Alelo A con respecto al genoma de referencia que están relacionados a la hora de producir una diferencia (marcada con el color correspondiente en la **Figura 10**) entre ambos alelos. La numeración está hecha según la posición en los correspondientes contigs del alelo (indicadas en verde para el contig01 y en naranja para el contig377box).



A continuación, se comparó la secuencia del ensamblaje de ambos alelos entre sí para detectar diferencias que puedan servir para el diseño de cebadores que permitan distinguir ambos alelos mediante PCR. Para ello, primero se sustrajo de cada alelo la región que corresponde al gen *BCNT2* según su alineamiento con el Gen 210 y el Gen 200 en el BLAST. En concreto, para el Alelo A de *BCNT2* se sustrajo desde el nt 98.633 hasta el nt 110.646 del contig01 y desde el nt 39.362 hasta el nt 42.144 del contig377box. Para el Alelo B de *BCNT2* se sustrajo desde el nt 27.377 hasta el nt 41.465 del contig04. Seguidamente se compararon ambas secuencias mediante BLAST (BLAST, 2018) (**Figura 10**). Entre las diferencias detectadas, cabe destacar la delección de 745 nts en el Alelo B con respecto al genoma de referencia y al Alelo A, que implica la pérdida del CDS2 del Gen 200.

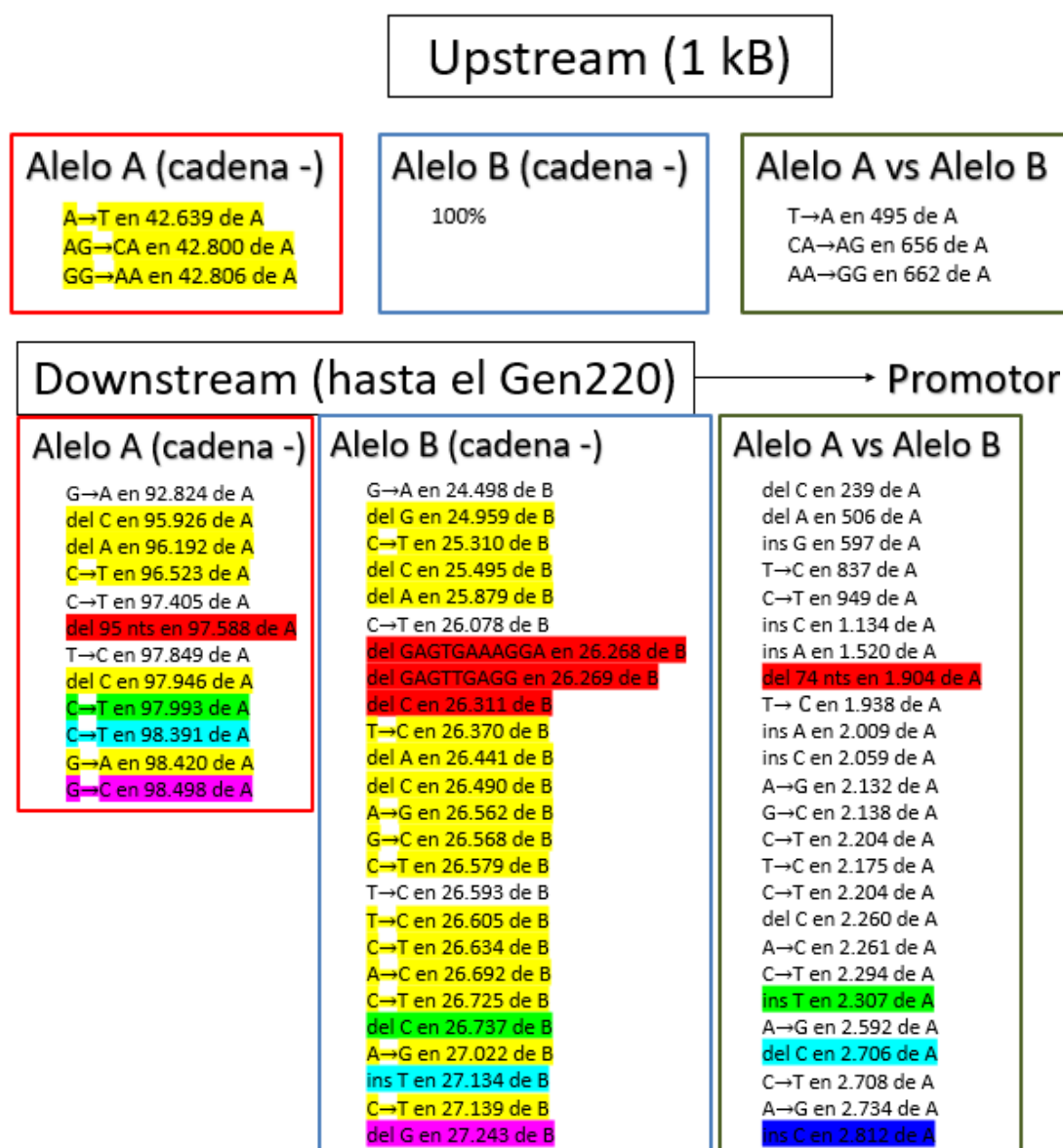
<b>Gen210 Int1:</b> C→T en 271 C→T en 275 T→A en 277 C→ en 429 del AGGGTTTT en 495  <b>Gen210 Int2:</b> C→A en 636 G→A en 744	<b>Gen210 Int3:</b> G→T en 927 G→A en 1.268 G→A en 1.364 del A en 1.564 A→G en 1.877 del TC en 1.993 del T en 2.010 del TT en 2.029 C→T en 2.205 C→T en 2.304 C→G en 2.387 A→G en 2.391 A→G en 2.496 G→A en 2.499 del AG en 2.686 C→T en 2.744 ins A en 2.836 C→T en 2.330 del T en 3.533 del CT en 3.593 T→C en 3.738 ins A en 3.789 T→A en 3.794 A→C en 4.078	<b>Gen210 Int4:</b> del T en 4.223 G→A en 4.269 ins TTT en 4.462  <b>Gen210 Int5:</b> del G en 4.567 del AC en 4.646 C→T en 5.369	<b>Gen200 UTR2:</b> C→T en 6.052  <b>Gen200 Int2+CDS2+Int3:</b> ins 745 nts en 6.474 de A  <b>Gen200 Int3:</b> del TC en 7.329 de A G→A en 7.518 de A del G en 7.621 de A del C en 7.776 de A del C en 7.780 de A  <b>Gen200 Int4:</b> T→C en 8.153 de A del AATC en 8.282 de A A→C en 8.405 de A G→A en 8.538 de A del G en 9.038 de A G→C en 9.070 de A C→T en 9.978 de A C→T en 10.010 de A del G en 10.153 de A del A en 10.262 de A A→G en 10.440 de A T→C en 10.451 de A C→T en 10.491 de A del TAAGCA en 10.507 de A del TC en 10.508 de A del G en 10.956 de A G→A en 10.957 de A T→C en 11.076 de A TG→CA en 11.078 de A del CACACACACA en 10.80 de A	del A en 11.092 de A del TTTA en 11.229 de A del G en 11.259 de A del T en 11.266 de A del T en 11.339 de A del G en 11.631 de A T→G en 11.140 de A T→C en 11.781 de A ins C en 11.783 de A G→T en 11.829 de A del T en 11.839 de A G→A en 12.066 de A C→T en 12.840 de A C→T en 12.963 de A  <b>Gen200 Int6:</b> ins AAA en 13.824 de A T→A en 13.976 de A A→G en 14.020 de A  <b>Gen200 UTR3:</b> ins G en 14.526 de A ins T en 14.532 de A G→C en 14.589 de A
--	---	---	---	---

**Figura 10.** Identificación de las diferencias entre los dos alelos de *BCNT2* de CS. En color aparecen los cambios que proceden de una combinación de varios polimorfismos de ambos alelos con respecto al genoma de referencia (marcados en fosforito con el color correspondiente en la **Figura 8** y en la **Figura 9**). La numeración está hecha según la posición en la secuencia de los dos contigs del Alelo A contiguos.

Finalmente, para comprobar que los polimorfismos detectados en regiones exónicas (que son los polimorfismos que pueden servir para comparar los niveles de transcripción de ambos alelos) no son errores procedentes de la de secuenciación de otros genes de la misma familia, comparamos el Gen 200 de cada alelo con la secuencia del genoma de referencia del locus *BCNT1* (genes *VIT\_05s0020g00080* y *VIT\_05s0020g00090*, a partir de ahora Gen 80 y Gen 90, respectivamente) mediante BLAST (BLAST, 2018). Los polimorfismos de las regiones UTR2 y CDS2 no pudieron ser analizados de esta manera porque la secuencia es indeterminada en la región del genoma de referencia de *BCNT1* que los albergaría. Por otro lado, los polimorfismos en el UTR3 del Gen 200 sí pudieron ser analizados, confirmándose que:

- El Alelo B de *VIT\_05s0020g00200* presenta una delección (del) de G que no coincide con una variación en esa posición entre los locus *BCNT1* y *BCNT2* en el genoma de referencia.
- El Alelo A de *VIT\_05s0020g00200* presenta una inserción (ins) de T que no coincide con una variación en esa posición entre los locus *BCNT1* y *BCNT2* en el genoma de referencia.
- El Alelo B de *VIT\_05s0020g00200* presenta una transversión G→C que no coincide con una variación en esa posición entre los locus *BCNT1* y *BCNT2* en el genoma de referencia.

Por otro lado, con el objetivo de identificar INDELs en las secuencias reguladoras que puedan explicar la sobreexpresión del alelo mutado extrajimos del genoma de referencia la secuencia del promotor (la secuencia aguas arriba del inicio de transcripción anotado para el gen *VIT\_05s0020g00210* que llega hasta la zona que comienza a estar anotada como gen *VIT\_05s0020g00220*) y, por otro lado, una secuencia de 1000 nts aguas abajo del final de la transcripción anotado para el gen *VIT\_05s0020g00200*. Estas se alinearon frente a los contigs del Alelo A y del Alelo B mediante BLAST (BLAST, 2018) (Figura 11).



**Figura 11.** Caracterización de la región upstream y downstream del Alelo A y del Alelo B de *BCNT2* en CS. En fosforito se presentan los polimorfismos que dan origen a una diferencia entre el Alelo A y el Alelo B. En fosforito con un color diferente al amarillo están anotados cada uno de los polimorfismos del Alelo A y del Alelo B con respecto al genoma de referencia que están relacionados a la hora de producir una diferencia (marcada con el color correspondiente) entre ambos alelos. La numeración está hecha según la posición en el contig del alelo que se especifica. Hay que tener en cuenta que como los genes están codificados en la cadena menos la región downstream es realmente el promotor.

Por último, recogimos todos los polimorfismos detectados (SNPs e INDELs) tanto en regiones codificantes, como en intrones, como en regiones reguladoras del Gen 210 y del Gen 200 para tener una visión global de las diferencias entre ambos alelos (**Tabla 3**). Cabe destacar los 2 SNPs y los 3 INDELs que aparecen en regiones codificantes del Gen 200 al comparar el Alelo A y el Alelo B.

**Tabla 3.** Resumen de los polimorfismos detectados en los alelos del gen *BCNT2* del parental CS. Aparecen los SNPs y los INDELs presentes tanto en regiones codificantes, como en intrones, como en regiones reguladoras del Gen 210 y del Gen 200.

### Resumen

Alelo A						
	Promotor	Gen210		Gen200		Downstream
		Intrones	Regiones codificantes	Intrones	Regiones codificantes	
SNPs	8	3	0	4	0	5
INDELs	4	12	0	18	1	0

Alelo B						
	Promotor	Gen210		Gen200		Downstream
		Intrones	Regiones codificantes	Intrones	Regiones codificantes	
SNPs	14	24	0	18	2	0
INDELs	11	5	0	9	2	0

Alelo A vs Alelo B						
	Promotor	Gen210		Gen200		Downstream
		Intrones	Regiones codificantes	Intrones	Regiones codificantes	
SNPs	13	23	0	22	2	5
INDELs	12	14	0	23	3	0

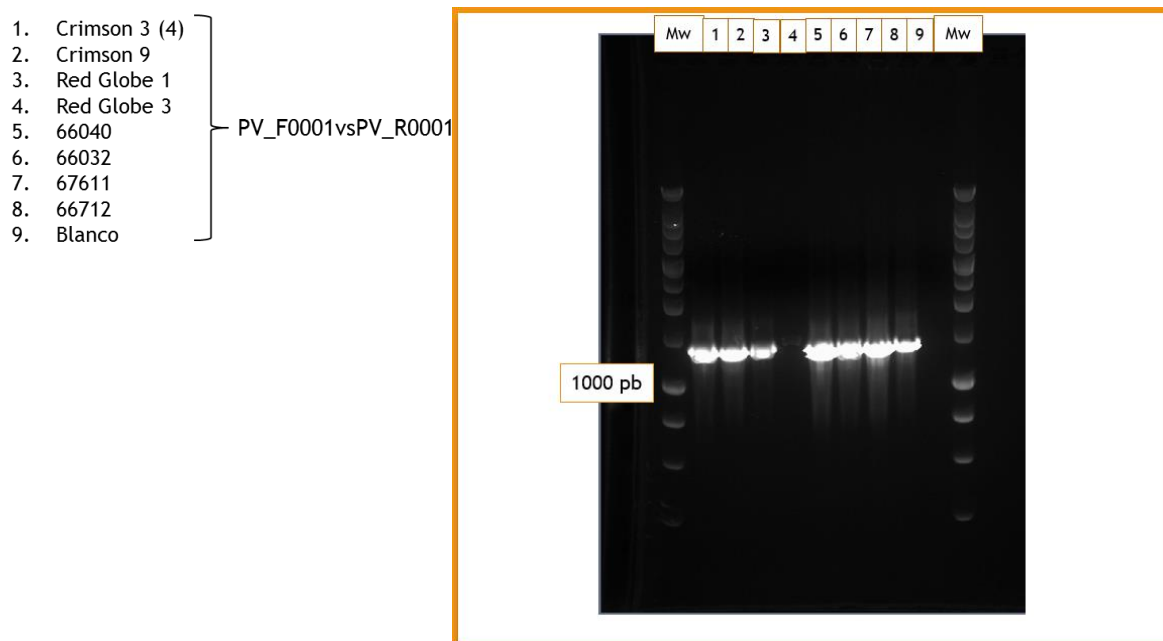
## 4.2. Confirmación por PCR seguida de secuenciación de los polimorfismos identificados en la región candidata.

### 4.2.1. Validación de los SNPs identificados en el UTR3 del gen *VIT\_05s0020g00200*.

En primer lugar, intentamos confirmar la presencia de los SNPs identificados en el UTR3 del gen *VIT\_05s0020g00200* (**Tabla 4**). Para ello, llevamos a cabo una PCR utilizando los cebadores PV\_F0001 y PV\_R0001 (**Figura 7 y Tablas 1 y 2**) sobre las extracciones de DNA de los parentales CS 3 (4), CS 9, RG 1 y RG 3 y de los individuos de la progenie F<sub>1</sub> RG X CS: 66040, 66032, 67611 y 66712 (**Figura 12**). Estos individuos F<sub>1</sub> fueron escogidos porque aparecen en las colas de la distribución fenotípica de forma de baya de la progenie. En concreto, 66040 y 66032 son individuos con baya elipsoidal y 67611 y 66712 son individuos con baya esférica y alta fertilidad (**Figura 5**).

**Tabla 4.** Resumen de los polimorfismos presentes en el UTR3 del Gen 200 y de las posiciones correspondientes.

Polimorfismo	Alelo A	Alelo B	Gen200_completo	Genoma de referencia	Secuenciación Alelo A	Secuenciación Alelo B
del C en B	41.872	41.195	1.272	5:2,078,172	14.526	13.819
ins A en A	41.878	41.201	1.266	5:2,078,166	14.532	13.824
C→G en B	41.915	41.236	1.230	5:2,078,130	14.569	13.860



**Figura 12.** Resultado de la PCR para intentar confirmar los polimorfismos del UTR3 del Gen 200.

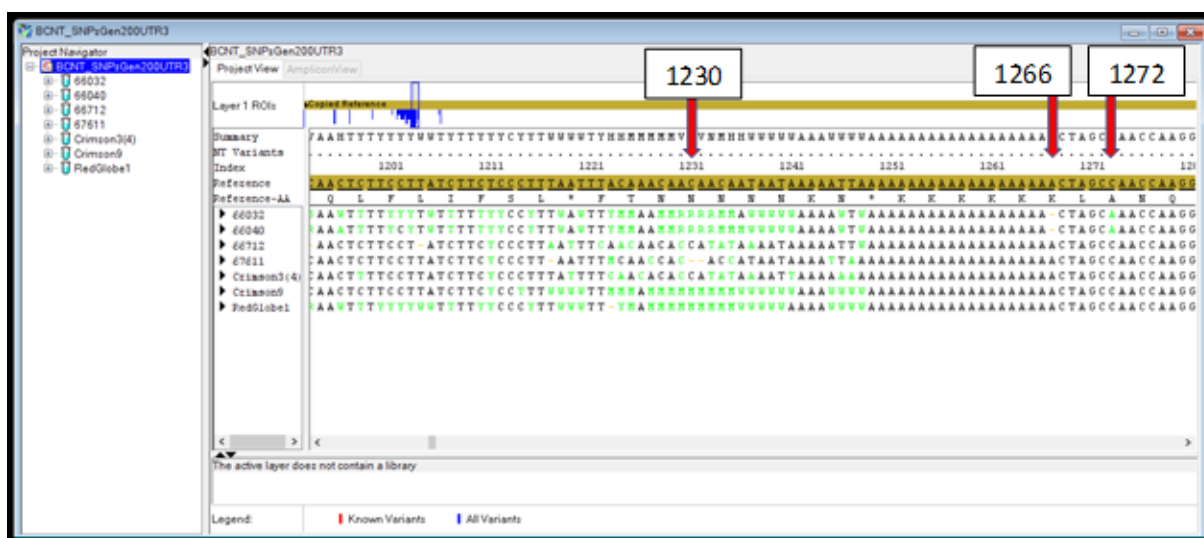
Dado los resultados positivos obtenidos en la PCR (**Figura 12**), decidimos enviar a secuenciar el producto de PCR de los individuos CS 3 (4), CS 9, RG 1, RG 3, 66040, 66032, 67611 y 66712 utilizando los mismos cebadores (PV\_F0001 y PV\_R0001) y analizamos las secuencias mediante SeqScape ('SeqScape v2.6', 2018).

En la posición 1.272, en la secuencia del amplicón obtenida, en vez de aparecer la delección de C esperada en el Alelo B se identificó un SNP con el cambio C→A en los individuos de la progenie con baya elipsoidal. Sin embargo, ni en CS ni en RG se identificó la A en la secuenciación. Los individuos 66032 y 66040 deben ser homocigotos por haber heredado el alelo con la A en la posición 1.272 de cada uno de los dos parentales, por tanto, la ausencia del alelo en los mismos podría explicarse porque en los parentales solo se haya amplificado el Alelo A. Las individuos de baya esférica y alta fertilidad 67611 y 66712 son homocigotos para C en esta posición, lo que indicaría que han heredado el Alelo A de ambos parentales.



Por otro lado, la ins A de la posición 1.266 predicha para el Alelo A aparece como una delección de A en los individuos 66032 y 66040 de la progenie. Este resultado indica de nuevo que estos individuos han heredado el Alelo B de CS y RG para esta posición (**Figura 13**). Por tanto, se deduce que el Alelo B es el alelo ligado a la sobreexpresión de *BCNT2* y a la mutación subyacente al QTL *FERTILITY/SHAPE* que determina el alargamiento de la baya y la disminución de la fertilidad.

Por último, el cambio C→G en el Alelo A de la posición 1.230 no puede ser analizado debido a que la presencia de los INDELs y repeticiones poli-A hace que se solapen los picos de la secuenciación de distintas moléculas en el cromatograma y no se pueda deducir la secuencia de cada alelo. Por otro lado, en un análisis RNAseq de otros individuos de la progenie disponible en el grupo (Royo *et al.*, 2018) sí que aparece el cambio C→G en esta posición. Sin embargo, en el individuo homocigoto para el Alelo B (66960) también aparece la G en el RNAseq, lo que sugiere que este polimorfismo podría no darse entre los alelos del Gen 200 como tal, sino que podría venir de la secuenciación del Gen 80 o de otras copias del Gen 200 en caso de que se hubiera duplicado en alguno de los alelos con respecto al genoma de referencia (**Figura 13**). Tras el análisis de esta posición también llegamos a la conclusión de que los individuos de baya elipsoidal presentan el Alelo B en homocigosis, que es el alelo que presentaría la sobreexpresión.



**Figura 13.** Resultados de la secuenciación de los SNPs del UTR3 del Gen 200.

Visto los resultados obtenidos, se preparó una PCR en gradiente para forzar la amplificación del Alelo B en los parentales a temperaturas de anillamiento de los cebadores más bajas, pero no se obtuvo amplificación por lo que es uno de los puntos que quedan por completar. Se asume que debió darse algún problema técnico en esta PCR puesto que había funcionado anteriormente (**Figura 12**).

Por otro lado, también se diseñó un nuevo cebador (PV\_F0012, **Tabla 2**) para intentar resolver la zona donde la presencia de los INDELs y poli-A hace que no se obtenga una secuencia definida. En esta secuenciación, a partir de las mismas amplificaciones de PCR descritas en la **Figura 13**, sí que se pudo analizar la posición 1.230 y se confirmó que en los individuos con baya elipsoidal aparece el cambio C→G. Sin embargo, ni en CS ni en RG aparece la G, lo que nuevamente nos lleva a pensar que en los parentales solo se ha amplificado el Alelo A en esa PCR.

#### *4.2.2. Validación de la delección del CDS2 y del SNP del UTR2 del gen VIT\_05s0020g00200.*

En segundo lugar, intentamos confirmar la presencia de la delección de 745 nt que resulta en la pérdida del CDS2 del Alelo B y del SNP del UTR2 del Gen 200 (G→A). Para ello, llevamos a cabo una PCR utilizando los cebadores PV\_F0006 y PV\_R0005 (**Figura 7 y Tablas 1 y 2**) sobre las extracciones de DNA de los parentales CS 3 (4), CS 9, RG 1 y de los individuos de la progenie 66040, 66032, 67611 y 66712 (**Figura 14**). La pareja de primers PV\_F0006-PV\_R0005 está diseñada de modo que los cebadores hibridan con el punto de rotura de la delección del CDS2 en el Alelo B, de manera que al obtener amplificación con esta pareja se confirma la presencia de la delección en el Alelo B. Por otro lado, el hecho de que no se obtenga amplificación en los individuos 67611 y 66712 indica que la delección no está en el Alelo A.

1. Crimson 9
  2. Red Globe 1
  3. 66040
  4. 66032
  5. 67611
  6. 66712
  7. Blanco
- } PV\_F0006vsPV\_R0005



**Figura 14.** Resultado de la PCR para intentar confirmar la delección del CDS2 y el SNP del UTR2 del Gen 200.

Como la PCR sobre la extracción del parental RG 1 no funcionó, repetimos la PCR con extracciones de DNA de varias muestras del parental RG (**Figura 15**).

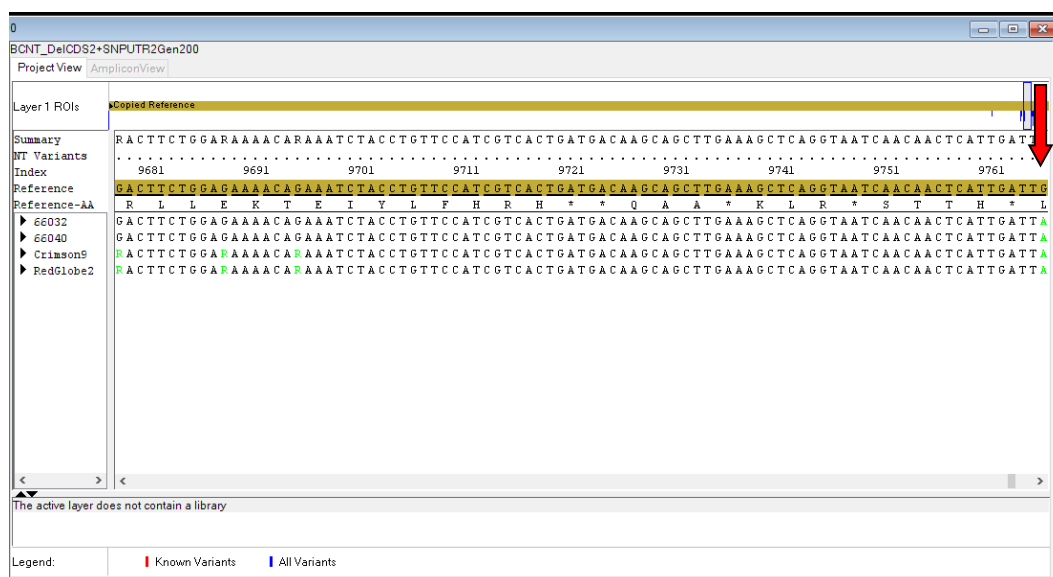
1. Red Globe 1
  2. Red Globe 2
  3. Red Globe 3
  4. Red Globe 4
  5. Red Globe 5
  6. Red Globe 6
  7. Red Globe 7
  8. Red Globe 8
  9. Red Globe 9
  10. Red Globe 10
  11. Crimson 9
  12. Blanco
- } PV\_F0006vsPV\_R0005



**Figura 15.** Resultado de la PCR para intentar confirmar la delección del CDS2 y el SNP del UTR2 del Gen 200 de los parentales RG.

Dado los resultados obtenidos en la PCR, decidimos enviar a secuenciar el producto de PCR de los individuos CS 9, RG 2, 66040 y 66032 utilizando los mismos cebadores (PV\_F0005 y PV\_R0005) y analizamos los datos mediante SeqScape ('SeqScape v2.6', 2018).

Con la secuenciación Sanger se pudo confirmar la presencia de la delección del exón y del SNP de la posición 9.765 (5:2,086,665 en el genoma de referencia) en los individuos 66040 y 66032. Al igual que para posiciones descritas anteriormente, este resultado confirma que estos individuos deben ser homocigóticos para el alelo B de la región. Sin embargo, en CS y en RG aparecen otros SNPs en otras posiciones del amplicón que no se habían identificado en el análisis de los contigs y que, por tanto, podrían venir del Gen 80 (para el que la secuencia en el genoma de referencia en esta región es indeterminada y no se puede comprobar) o de otra copia del Gen 200 que también tendría la delección (**Figura 16**). Estos SNPs no aparecen en los datos de RNAseq.



**Figura 16.** Resultados de la secuenciación de la delección del CDS2 y del SNP del UTR2 del Gen 200.

### 4.3. Análisis de la secuencia de los transcritos de los genes *BCNT* en los datos de RNAseq.

En tercer lugar, para seguir profundizando en el conocimiento de la secuencia y de la expresión de estos alelos, decidimos comprobar la presencia de los polimorfismos identificados en la secuenciación de PacBio en datos de secuenciación RNAseq de transcritos de individuos de la progenie F<sub>1</sub> RG X CS. Estos datos RNA-seq fueron producidos en trabajos anteriores del grupo (Royo *et al.*, 2018) para los individuos de la progenie 66960, 67017, 67517, 66517, 67119 y 67706 (**Tabla 4**). En este trabajo de máster se analizó la expresión de los genes *BCNT* en la progenie revisando el alineamiento de las secuencias RNAseq frente al genoma de referencia mediante el visualizador IGV ('IGV\_2.4.8', 2018).

En el individuo 66960, que según el genotipado de marcadores moleculares realizado para el análisis de QTLs debe ser homocigoto para el Alelo B asociado al desarrollo de bayas elipsoidales (**Tabla 5**), aparecen *reads* de RNAseq que confirman la delección del CDS2 del Alelo B en algunos transcritos, pero en otros *reads* del mismo individuo también aparece el exón (**Figura 16A**), lo que podría indicar que viene de transcritos del Gen 80 o de otras copias del Gen 200. En los individuos 67017, 67517 y 66517 vemos ausencia de expresión, lo que podría correlacionar con la presencia del Alelo A en homocigosis. Esto confirmaría que la sobreexpresión de *BCNT* ocurre asociada a la presencia del Alelo B en los individuos F<sub>1</sub>. Por último, en los individuos 67119 y 67706, que en principio son heterocigotos con Alelo A y B (**Tabla 5**), hemos detectado el SNP de la posición 5:2,078,130 que corresponde al Alelo B (**Figura 16B**), pero también aparecen *reads* iguales al genoma de referencia, lo que podría indicar que vienen del Gen 80, del Alelo A o de otras copias del Gen 200. Además, se detectó el SNP 5:2,086,665 que corresponde al Alelo B (**Figura 16C**), lo que apoyaría la hipótesis de que solo se expresa el Alelo B del Gen 200 y que los otros *reads* vienen del Gen 80 o de otras copias del Gen 200 ligadas en el misma copia cromosómica. En estos individuos de genotipo A/B también aparecen *reads* que apoyan la existencia de la delección predicha para el Alelo B, pero en otros *reads* aparece el exón, lo

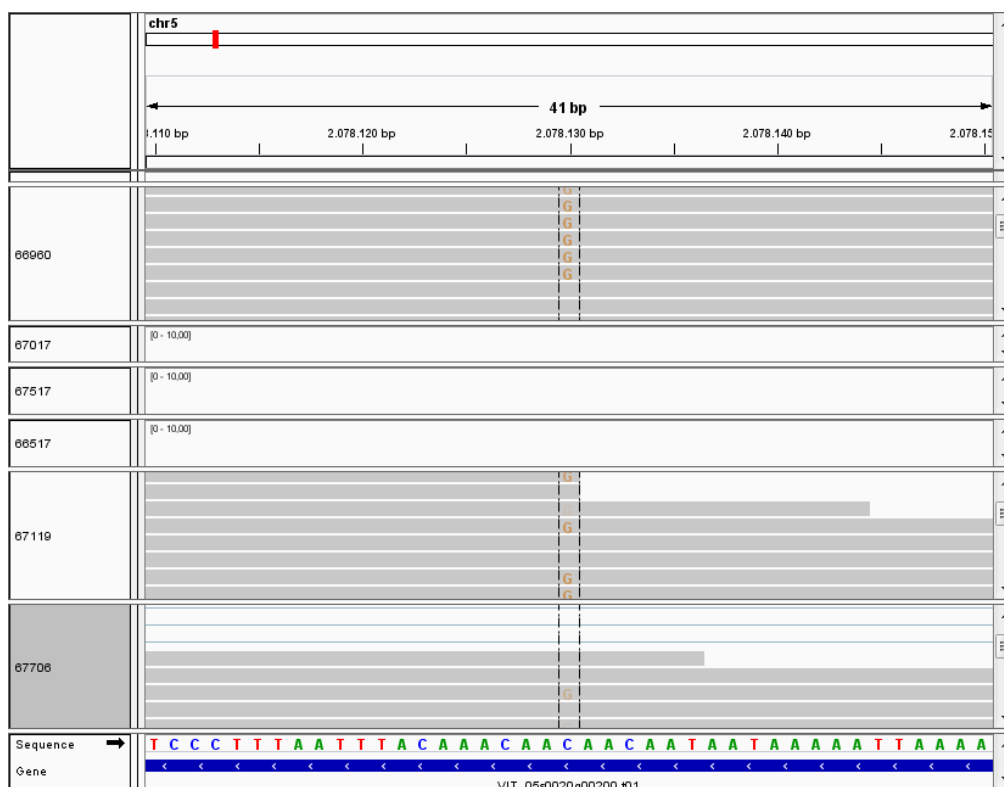
que podría indicar que vienen del Gen 80 o de otras copias del Gen 200 ligadas en el haplotipo B como se ha descrito para el individuo 66960 homocigoto para el Alelo B.

**Tabla 5.** Datos de los individuos de la progenie sobre los que se llevó a cabo el ensayo de RNAseq.

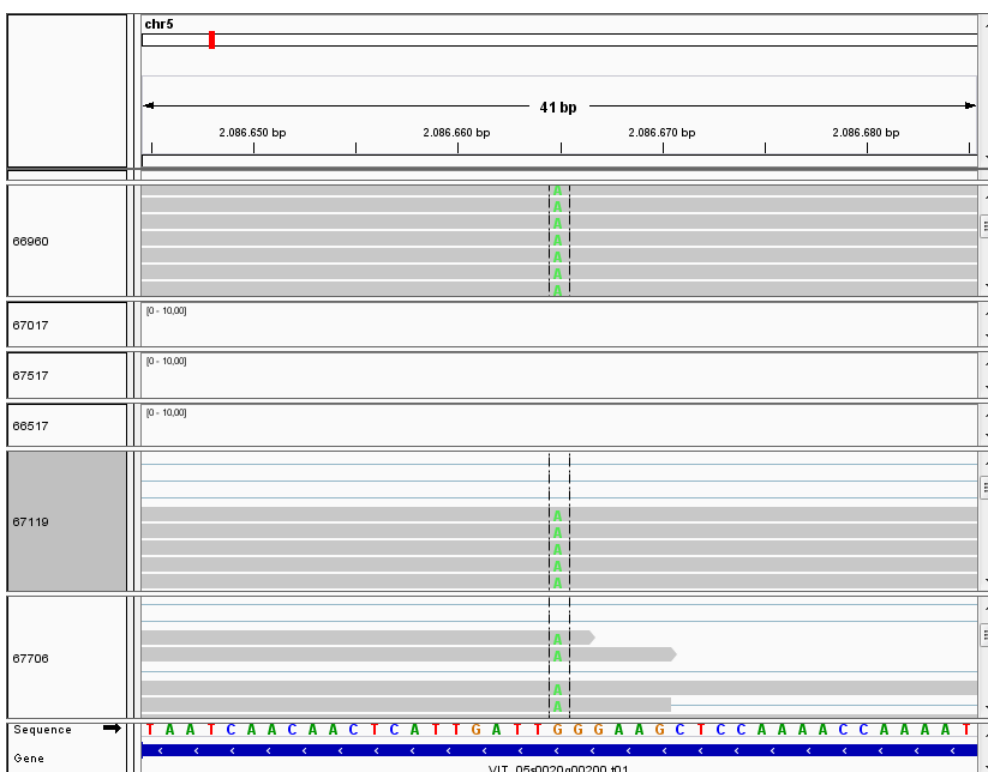
	Genotipo (Royo <i>et al.</i> , 2018)	Fenotipo de forma de baya asociado	Índice de forma de baya	Predicción de los alelos
66960	kk nn	Elipsoidal	1'13	B/B
67017			1'05	
67517	hh nn	Esférica/Elipsoidal	1'07	A/A
66517	hh np	Esférica	1'18	A/A o A/B
67119	hk nn	Elipsoidal	1'09	A/B
67706	hk np	Esférica	1'24	A/B



**Figura 16A.** Confirmación de la delección del CDS2 del Alelo B en los datos de RNAseq.



**Figura 16B.** Confirmación del SNP de la posición 5:2,078,130 del Alelo B en los datos de RNA seq.



**Figura 16C.** Confirmación del SNP de la posición 5:2,086,665 del Alelo B en los datos de RNA seq.

Visto los resultados obtenidos, decidimos llevar a cabo un experimento de qPCR para cuantificar el número de copias (Gaines *et al.*, 2010) de los genes de la familia BCNT con el objetivo de discernir si los SNPs que no habían sido detectados previamente en la caracterización de los alelos proceden del Gen 80 o de otras copias del Gen 200 que hayan podido aparecer en el Alelo B al menos.

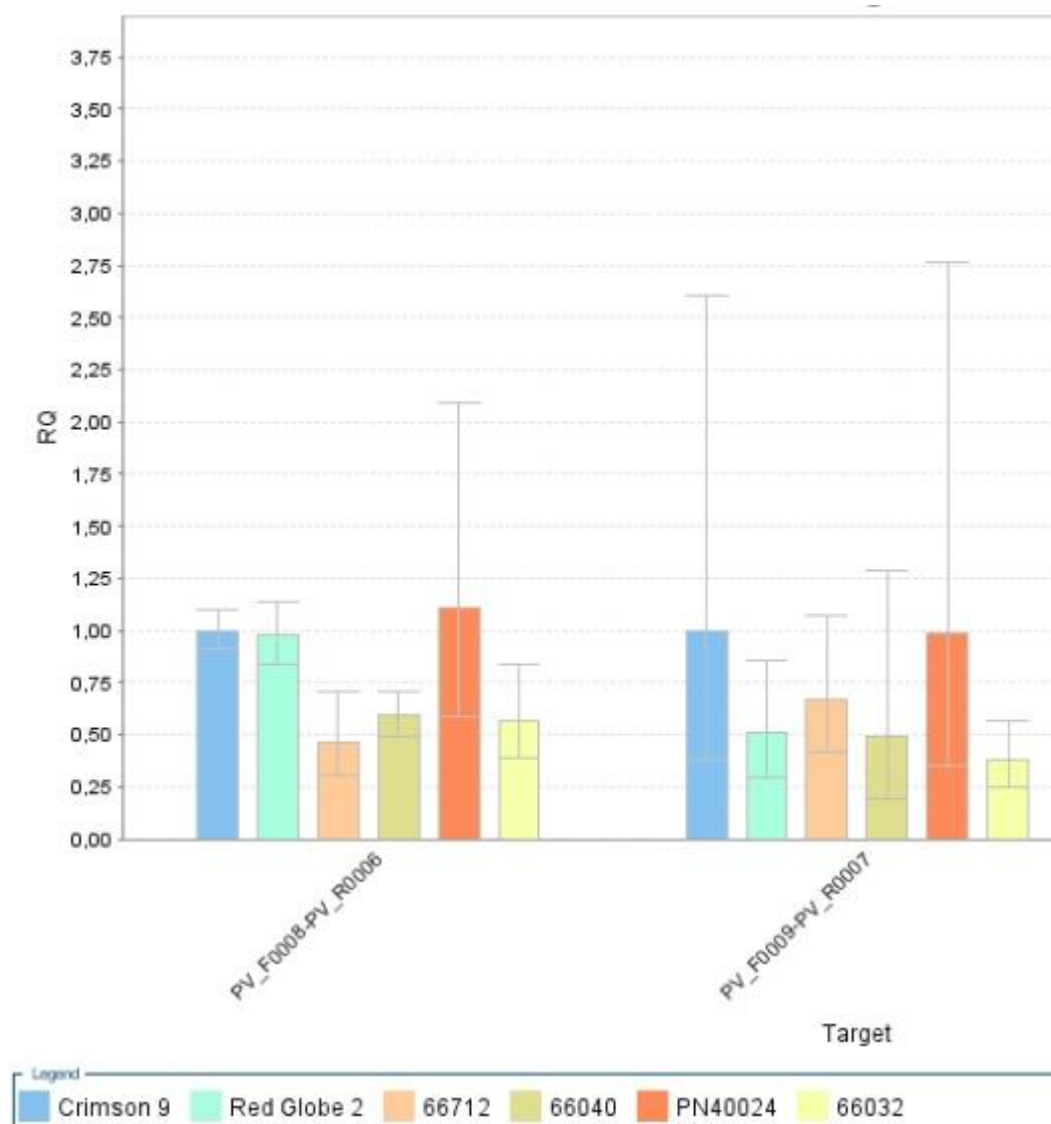
#### **4.4. Cuantificación del número de copias de los genes de la familia BCNT.**

Con el objetivo de discernir si los SNPs que no habían sido previamente detectados en la caracterización de los alelos proceden del Gen 80 o de otras copias del Gen 200 decidimos llevar a cabo un experimento de qPCR. En este experimento los niveles de expresión fueron determinados en función del número de ciclos de amplificación necesarios para alcanzar el umbral fijado en la fase exponencial de la reacción de amplificación ( $C_t$ ). Por otro lado, los niveles relativos fueron normalizados a la cantidad de DNA genómico presente en cada muestra en función de los niveles de amplificación obtenido para el gen *LEAFY* (VIT\_17s0000g00150), el cual está presente en una sola copia en el genoma de la vid (Carmona, Cubas and Martinez-Zapater, 2002).

Para el análisis se diseñaron dos parejas de cebadores (**Tabla 2**) que permiten la amplificación de los tres miembros de la familia BCNT descritos en el genoma de referencia de la vid (**Figura 6**). Como referencia de la relación tres a uno esperada para el número de copias entre *BCNT* y *LEAFY*, se utilizó en las amplificaciones DNA de PN40024 que es el genotipo correspondiente al genoma de referencia sobre el que hemos estimado esta relación. Los resultados (**Figura 17**) no mostraron diferencias significativas en el número relativo de copias de los genes *BCNT* frente al gen *LEAFY* para ninguna de las dos parejas de cebadores empleadas para los genes *BCNT* (PV\_F0008-PV\_R0006 y PV\_F0009-PV\_R0007) y para ninguno de los individuos. El análisis incluyó a los parentales de la progenie RG y CS, que según los análisis de secuenciación descritos anteriormente deben ser heterocigóticos A/B para la región en la que se codifican los genes *BCNT*. Dado que la relación del número de copias obtenido en CS y RG no es diferente a la obtenida en PN40024



(Figura 17), asumimos que tanto en el alelo A como en el B se mantienen las tres copias de *BCNT* descritas en el genoma de referencia y que no se han producido duplicaciones de ningún *BNCT* ni en el alelo A ni en el B. De hecho, con la pareja de cebadores PV\_F0008-PV\_R0006, aunque la diferencia no sea significativa, el número de copias en los individuos de la progenie homocigóticos para el alelo B (66040 y 66032) tiende a ser menor que en los parentales, lo cual iría en contra de la idea de que el aumento de expresión de *BCNT* en individuos que presentan el Alelo B se debe a la duplicación de algún miembro de la familia en la copia cromosómica ligada al Alelo B. En conjunto, los resultados de la qPCR parecen indicar que los SNPs que no habían sido detectados previamente en la secuenciación del Gen 200 proceden del Gen 80 y no de otras copias del Gen 200 derivadas de la duplicación del original. Es decir, a pesar de que los cebadores fueron diseñados específicamente para el Gen 200, los resultados muestran que en los ensayos de secuenciación puede aparecer amplificación inespecífica del Gen 80. Sin embargo, teniendo en cuenta los resultados del número de copias y de la secuenciación de RNAseq, sí parece que en los individuos con baya elipsoidal existe una sobreexpresión de lo que hemos caracterizado como Alelo B del Gen 200, pero también una sobreexpresión de uno de los alelos del Gen 80. Este resultado podría explicarse con la presencia de una mutación en el Alelo B que causase la sobreexpresión en *cis* de ambos genes pese a estar separados por más de 100 kb en el genoma (Figura 6). Para confirmar esto será necesario analizar las regiones reguladoras de ambos genes en trabajos posteriores, así como la realización de qRT-PCR alelo-específica en ambos genes para determinar si efectivamente, el Alelo B de los genes 80 y 200 está sobreexpresado en individuos heterocigóticos A/B.



**Figura 17.** Resultados del experimento de qPCR para cuantificar el número de copias de los genes de la familia BCNT. Se muestra la relación (RQ) entre los genes BCNT y el gen LEAFY en los individuos CS 9, RG 2, 66040, 66032, 66712 y PN40024. Se muestran los resultados obtenidos tras emplear dos parejas de cebadores distintas para la amplificación de los genes *BCNT*.

## 5. CONCLUSIONES.

1. Los contigs derivados de la secuenciación PacBio de BACs de CS corresponden a dos alelos diferentes del gen *BCNT2*.
2. La combinación de resultados de secuenciación dirigida y de RNAseq confirma que la sobreexpresión de genes *BCNT* está ligada a la presencia del alelo B de *BCNT2* y que tanto *BCNT1* como *BCNT2* podrían estar sobreexpresados en los individuos que presentan dicho alelo.
3. La sobreexpresión de genes *BCNT* ligada a la presencia del alelo B no se relaciona con duplicaciones de ningún miembro de la familia.
4. Se detectaron deleciones de gran tamaño específicas del alelo B en posibles regiones reguladoras de *BCNT2* que podrían relacionarse con la sobreexpresión de los miembros de la familia. Se requieren más análisis moleculares y funcionales para comprobar si estas u otras mutaciones presentes en el alelo B podrían causar la sobreexpresión en *cis* de *BCNT1* y *BCNT2* y alterar con ello la determinación floral de las yemas y la morfogénesis del fruto.

## 6. BIBLIOGRAFÍA

- Altschul, S. F. *et al.* (1990) 'Basic local alignment search tool', *Journal of molecular biology*. Elsevier, 215(3), pp. 403–410.
- Baldet, P. *et al.* (2006) 'The expression of cell proliferation-related genes in early developing flowers is affected by a fruit load reduction in tomato plants', *Journal of Experimental Botany*. Oxford University Press, 57(4), pp. 961–970.
- Barrero, L. S. and Tanksley, S. D. (2004) 'Evaluating the genetic basis of multiple-locule fruit in a broad cross section of tomato cultivars', *Theoretical and Applied Genetics*. Springer, 109(3), pp. 669–679.
- Bertin, N. *et al.* (2003) 'Do Genetic Make-up and Growth Manipulation Affect Tomato Fruit Size by Cell Number, or Cell Size and DNA Endoreduplication?', *Annals of Botany*. Oxford University Press, 92(3), pp. 415–424.
- BLAST (2018). Available at:  
[https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=MegaBlast&PROGRAM=blastn&BLAST\\_PROGRAMS=megaBlast&PAGE\\_TYPE=BlastSearch&BLAST\\_SPEC=blast2seq&DATABASE=n/a&QUERY=&SUBJECTS=](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=MegaBlast&PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&DATABASE=n/a&QUERY=&SUBJECTS=).
- BLAT (2018). Available at: <http://www.genoscope.cns.fr/blat-server/cgi-bin/vitis/webBlat>.
- Borzecka, E. (2014) *The genetic and molecular bases of berry shape in grapevine (Vitis Vinifera L.)*.
- Cabezas, J. A. *et al.* (2006) 'A genetic analysis of seed and berry weight in grapevine', *Genome*. NRC Research Press, 49(12), pp. 1572–1585.
- Carmona, M. J. *et al.* (2008) 'A molecular genetic perspective of reproductive development in grapevine', *Journal of experimental botany*. Oxford University Press, 59(10), pp. 2579–2596.
- Carmona, M. J., Cubas, P. and Martinez-Zapater, J. M. (2002) 'VFL, the grapevine FLORICAULA/LEAFY ortholog, is expressed in meristematic regions independently of their fate', *Plant Physiology*. Am Soc Plant Biol, 130(1), pp. 68–77.

- Cheniclet, C. *et al.* (2005) 'Cell expansion and endoreduplication show a large genetic variability in pericarp and contribute strongly to tomato fruit growth', *Plant Physiology*. Am Soc Plant Biol, 139(4), pp. 1984–1994.
- Conde, C. *et al.* (2007) 'Biochemical changes throughout grape berry development and fruit and wine quality'. Global Science Books.
- Cong, B., Barrero, L. S. and Tanksley, S. D. (2008) 'Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication', *Nature genetics*. Nature Publishing Group, 40(6), p. 800.
- Considine, xJ A. and Knox, R. B. (1979) 'Development and histochemistry of the pistil of the grape, *Vitis vinifera*', *Annals of Botany*. Oxford University Press, 43(1), pp. 11–22.
- Coombe, B. G. (1960) 'Relationship of growth and development to changes in sugars, auxins, and gibberellins in fruit of seeded and seedless varieties of *Vitis vinifera*', *Plant physiology*. American Society of Plant Biologists, 35(2), p. 241.
- Coombe, B. G. (1972) 'The regulation of set and development of the grape berry', in *Symposium on growth Regulators in Fruit Production 34*, pp. 261–274.
- Coombe, B. G. (1976) 'The development of fleshy fruits', *Annual Review of Plant Physiology*. Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, 27(1), pp. 207–228.
- Le Cunff, L. *et al.* (2008) 'Construction of nested genetic core collections to optimize the exploitation of natural diversity in *Vitis vinifera* L. subsp. *sativa*', *BMC Plant Biology*. BioMed Central, 8(1), p. 31.
- Dai, Z. W. *et al.* (2011) 'Ecophysiological, genetic, and molecular causes of variation in grape berry weight and composition: a review', *American Journal of Enology and Viticulture*. Am Soc Enol Viticulture, 62(4), pp. 413–425.
- Dauelsberg, P. *et al.* (2011) 'Effect of pollination and fertilization on the expression of genes related to floral transition, hormone synthesis and berry development in grapevine', *Journal of plant physiology*. Elsevier, 168(14), pp. 1667–1674.

Deluc, L. G. *et al.* (2007) 'Transcriptomic and metabolite analyses of Cabernet Sauvignon grape berry development', *BMC genomics*. BioMed Central, 8(1), p. 429.

*DNA extraction* (2018). Available at: <https://www.qiagen.com/us/shop/sample-technologies/dna/genomic-dna/dneasy-plant-mini-kit/>.

Dokoozlian, N. K. (2000) 'Grape berry growth and development', *Raisin production manual*. University of California, Agricultural and Natural Resources Publication 3393 Oakland, CA, 3393, p. 30.

Doligez, A. *et al.* (2002) 'Genetic mapping of grapevine (*Vitis vinifera* L.) applied to the detection of QTLs for seedlessness and berry weight', *Theoretical and Applied Genetics*. Springer, 105(5), pp. 780–795.

Ebadi, A., Moghadam, J. E. and Fatahi, R. (2009) 'Evaluation of 22 populations achieved from controlled crossing between some seededx seedless grapevine cultivars', *Scientia horticulturae*. Elsevier, 119(4), pp. 371–376.

*ExoSAP-IT* (2018). Available at:  
<http://www.thermofisher.com/order/catalog/product/78201.1.ML>.

Fernandez, L. *et al.* (2013) 'Mis-expression of a PISTILLATA-like MADS box gene prevents fruit development in grapevine', *The Plant Journal*. Wiley Online Library, 73(6), pp. 918–928.

Fournier-Level, A. *et al.* (2009) 'Quantitative genetic bases of anthocyanin variation in grape (*Vitis vinifera* L. ssp. *sativa*) berry: a quantitative trait locus to quantitative trait nucleotide integrated study', *Genetics*. Genetics Soc America, 183(3), pp. 1127–1139.

Gaines, T. A. *et al.* (2010) 'Gene amplification confers glyphosate resistance in *Amaranthus palmeri*', *Proceedings of the National Academy of Sciences*. National Acad Sciences, 107(3), pp. 1029–1034.

Gepts, P. (2010) 'Crop domestication as a long-term selection experiment', *Plant breeding reviews*, 24(Part 2), pp. 1–44.

Gillaspy, G., Ben-David, H. and Gruissem, W. (1993) 'Fruits: a developmental perspective.', *The Plant Cell*. American Society of Plant Biologists, 5(10), p. 1439.

Grandillo, S., Ku, H. M. and Tanksley, S. D. (1999) 'Identifying the loci responsible for natural variation in fruit size and shape in tomato', *Theoretical and Applied Genetics*. Springer, 99(6), pp. 978–987.

Grape BLAST (2018). Available at:  
<http://genomes.cribi.unipd.it/grape/blast/blast.php>.

Hackbusch, J. *et al.* (2005) 'A central role of Arabidopsis thaliana ovate family proteins in networking and subcellular localization of 3-aa loop extension homeodomain proteins', *Proceedings of the National Academy of Sciences of the United States of America*. National Acad Sciences, 102(13), pp. 4908–4912.

Houel, C. *et al.* (2013) 'Genetic variability of berry size in the grapevine (*Vitis vinifera* L.)', *Australian journal of grape and wine research*. Wiley Online Library, 19(2), pp. 208–220.

'IGV\_2.4.8' (2018).

Jaillon, O. *et al.* (2007) 'The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla', *nature*. Nature Publishing Group, 449(7161), p. 463.

Keller, M. (2015) *The science of grapevines: anatomy and physiology*. Academic Press.

Kent, W. J. (2002) 'BLAT—the BLAST-like alignment tool', *Genome research*. Cold Spring Harbor Lab, 12(4), pp. 656–664.

Van der Knaap, E. and Tanksley, S. D. (2003) 'The making of a bell pepper-shaped tomato fruit: identification of loci controlling fruit morphology in Yellow Stuffer tomato', *Theoretical and Applied Genetics*. Springer, 107(1), pp. 139–147.

Kuulasmaa, T., Javed, A. A. and Muthusamy, A. (2004) 'Oligo Explorer 1.2', *Genelink*, Hawthorne, New York.

- Liu, J. *et al.* (2002) 'A new class of regulatory genes underlying the cause of pear-shaped tomato fruit', *Proceedings of the National Academy of Sciences*. National Acad Sciences, 99(20), pp. 13302–13306.
- Lozano, R. *et al.* (2009) 'Genetic analysis of reproductive development in tomato', *International Journal of Developmental Biology*. UPV/EHU Press, 53(8-9–10), pp. 1635–1648.
- Mejía, N. *et al.* (2007) 'Identification of QTLs for seedlessness, berry size, and ripening date in a seedless x seedless table grape progeny', *American journal of enology and viticulture*. Am Soc Enol Viticulture, 58(4), pp. 499–507.
- Messina, G. *et al.* (2014) 'Yeti, an essential *Drosophila melanogaster* gene, encodes a protein required for chromatin organization', *J Cell Sci*. The Company of Biologists Ltd, 127(11), pp. 2577–2588.
- Messina, G. *et al.* (2015) 'The Bucentaur (BCNT) protein family: a long-neglected class of essential proteins required for chromatin/chromosome organization and function', *Chromosoma*. Springer, 124(2), pp. 153–162.
- Messina, G. *et al.* (2016) 'Expression of human Cfdp1 gene in *Drosophila* reveals new insights into the function of the evolutionarily conserved BCNT protein family', *Scientific reports*. Nature Publishing Group, 6, p. 25511.
- Messina, G. *et al.* (2017) 'The human Cranio Facial Development Protein 1 (Cfdp1) gene encodes a protein required for the maintenance of higher-order chromatin organization', *Scientific reports*. Nature Publishing Group, 7, p. 45022.
- Monforte, A. J. *et al.* (2014) 'The genetic basis of fruit morphology in horticultural crops: lessons from tomato and melon', *Journal of experimental botany*. Oxford University Press, 65(16), pp. 4625–4637.
- Nanodrop (2018). Available at:  
<https://www.thermofisher.com/us/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/ultraviolet-visible-visible-spectrophotometry-uv-vis-vis/uv-vis-vis-instruments/nanodrop-microvolume-spectrophotometers.html>.



OIV (2018). Available at: <http://www.oiv.int/public/medias/5479/oiv-en-bilan-2017.pdf>.

Ojeda, H. *et al.* (1999) 'Berry development of grapevines: relations between the growth of berries and their DNA content indicate cell multiplication and enlargement', *Vitis*, 38(4), pp. 145–150.

Ollat, N. *et al.* (2002) 'Grape berry development: a review', *International Journal of Vine and Wine Sciences*.

Ozga, J. A. and Reinecke, D. M. (2003) 'Hormonal interactions in fruit development', *Journal of Plant Growth Regulation*. Springer, 22(1), pp. 73–81.

Paris, H. S. *et al.* (2003) 'Assessment of genetic relationships in Cucurbita pepo (Cucurbitaceae) using DNA markers', *Theoretical and Applied Genetics*. Springer, 106(6), pp. 971–978.

'PCR' (2018).

Périn, C. *et al.* (2002) 'Genetic control of fruit shape acts prior to anthesis in melon (*Cucumis melo* L.)', *Molecular Genetics and Genomics*. Springer, 266(6), pp. 933–941.

Pratt, C. (1971) 'Reproductive anatomy in cultivated grapes-a review', *American Journal of Enology and Viticulture*. Am Soc Enol Viticulture, 22(2), pp. 92–109.

*qPCR* (2018). Available at:

<http://www.thermofisher.com/order/catalog/product/4367659>.

Rodríguez, G. R. *et al.* (2011) 'Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity', *Plant physiology*. Am Soc Plant Biol, 156(1), pp. 275–285.

Rosati, A. *et al.* (2011) 'Differences in ovary size among olive (*Olea europaea* L.) cultivars are mainly related to cell number, not to cell size', *Scientia horticulturae*. Elsevier, 130(1), pp. 185–190.

Royo, C. *et al.* (2018) 'The major origin of seedless grapes is associated with a missense mutation in the MADS-box gene *VviAGL11*', *Plant physiology*. Am Soc Plant Biol, p. pp-00259.

Sánchez, L. A. and Dokoozlian, N. K. (2005) 'Bud microclimate and fruitfulness in *Vitis vinifera* L.', *American Journal of Enology and Viticulture*. Am Soc Enol Viticulture, 56(4), pp. 319–329.

*Secuenciación Sanger* (2018). Available at:  
<https://www.stabvida.com/es/servicio-de-secuenciacion-sanger>.

'SeqScape v2.6' (2018).

Spjut, R. W. (1994) *A systematic treatment of fruit types*. New York Botanical Garden.

Sun, L. and Luk, E. (2017) 'Dual function of Swc5 in SWR remodeling ATPase activation and histone H2A eviction', *Nucleic acids research*. Oxford University Press, 45(17), pp. 9931–9946.

Tanksley, S. D. (2004) 'The genetic, developmental, and molecular bases of fruit size and shape variation in tomato', *The plant cell*. Am Soc Plant Biol, 16(suppl 1), pp. S181–S189.

Vasconcelos, M. C. *et al.* (2009) 'The flowering process of *Vitis vinifera*: a review', *American Journal of Enology and Viticulture*. Am Soc Enol Viticulture, 60(4), pp. 411–434.

Zepeda, M. *et al.* (2006) 'Effects of the application of cytokinin bioregulators on inflorescence development of table grapes', *Acta horticulturae*.

## 7. ANEXOS

### 7.1. Abreviaturas:

BACs: Bacterial Artificial Chromosomes.

BCNT: Bucentaurus.

CDSs: secuencias codificantes.

CS: Crimson Seedles.

Del: delección.

DNA: ácido desoxirribonucleico.

FT: Factor de Transcripción.

Ins: inserción.

Int: intrones.

LG: grupos de ligamiento.

Nt: nucleótido.

PCR: Reacción en Cadena de la Polimerasa.

OFP: Ovate Family Protein.

QTL: Quantitative Trait Locus.

QTLs: Quantitative Trait Loci.

RG: Red Globe.

SNPs: single nucleotide polymorphisms.

UTRs: untraslated regions.

